A STUDY OF VIGILANCE AND FATIGUE

BY

D.C. FRASER

FOREWORD

This study of Vigilance and Fatigue in relation to Breakdown of Skilled Performance was carried out from 1952 to 1956. All the experimental work described was originated and carried out by the writer, and none of this work has been used as part of another thesis.

It is a pleasure to express my thanks to Sir Frederic Bartlett for advice and discussion; to Professor Bradford Hill for advice and discussion on certain statistical points; to Dr. J.S. Weiner and Mr. A.R. Lind, of the Department of Human Anatomy, University of Oxford, for discussion of the work on heat stress; to Mr. D.G.C. Gronow for help with the statistics; and to Professor Drever for supervising the research.

INTRODUCTION

The subject of this study is based on a concept first discussed by Sir Henry Head. In 1926 he postulated a state of what he called "neural vigilance" - a condition of maximum physiological efficiency. The determinants of this condition were never very clearly defined or isolated, and on the physiological side it remains little more than one of those attractive speculations which enliven the humdrum routine of research into the complications of human functioning. On the psychological side, however, the concept has been taken up and developed in some detail, first of all by Dr. N. H. Mackworth working at the Psychological Laboratory in Cambridge; then by the writer and by Mr. D. E. Broadbent, working in the same department; and recently by certain research workers on the other side of the Atlantic. The present study deals largely with the applications of certain techniques, originally devised in the laboratory, to field situations of several kinds. As will be seen, the rather narrow scope of the earlier investigation broadened out considerably, and in the end overlaps two other important fields of psychological research.

The main theme of the work can be summed up under three headings:

Previous experimental work has almost invariably studied changes in vigilance as affected by the present situation - i.e. the effect of, say, duration of watch on vigilance during the watch itself, or the effect of noise occurring during the performance of the vigilance task. But it is a major tenet of this study that, if vigilance is more than just a convenient label for a certain group of situations, it should be possible to demonstrate changes in vigilance occurring as the result of conditions extraneous to the vigilance task and not necessarily contemporaneous with it.

The initial set of experiments, accordingly, was designed to test the truth of this hypothesis.

This being established, it was clearly desirable to examine the pattern of changes in vigilance in detail. But, at this point in the work, the writer developed a theory to account for the fact that vigilance tasks appeared to be giving positive results in situations, e.g. "fatigue" and ambient noise, which had yielded little or no results in previous research. Much of the next portion of the experimental work, therefore, is concerned with testing the soundness of this theory as well as in examining in more detail the pattern of changes in vigilance as the result of certain forms of "fatigue".

Lastly, since fatigue is regarded as closely allied to stress, an examination is made of the changes in vigilance as the result of exposure to a rather different form of stress. Certain predictions were made about changes in the form of the vigilance test to make it more sensitive in certain respects, and these predictions are then tested in the final section while additional data were accumulated about the pattern of changes in vigilance under this other form of stress.

The study sets out, then, to do three things:

(a) To test three sets of broad hypotheses about the nature of vigilance and its relation to fatigue and stress.

(b) To collect a substantial amount of information about the pattern of changes in vigilance under certain forms of stress situation.

(c) To discuss the relation of this information to other work in the field of experimental psychology and physiology.

PREVIOUS RESEARCH ON VIGILANCE

The literature in the field of vigilance is fairly quickly covered, since the total amount of work done is small enough for the present quota to constitute a substantial proportion of the total. Perhaps one of the earliest experiments related to this work, although not specifically using the concept, is some work of Wyatt and Langdon (1932) on cartridge case inspection. Here they found a marked decrease in the number of rejects after the first 30 to 45 minutes. This decrease continued for some 90 minutes, and was followed by an irregular recovery.

A rather drastic experiment was carried out by Lindsley et al (1944). In this eight practised subjects operated radar sets continuously for 4 hours per day, 6 days per week for 17 days. The rate of presentation in this experiment was high - 116 signals in each 4-hour session. They found a significant decrease in the second, third and fourth hours compared with the first, but this decrease did not appear until the third day.

The concept of vigilance, however, was first introduced as a major theoretical concept by Mackworth. The experiments described in his book (Mackworth, 1950) began with some observations on the number of submarines detected by Coastal Command aircraft on anti-submarine patrol, and extended over some six years. Much of his work was carried out with a Clock Test, which he devised to present what he held to be the essential features of the vigilance situation - a regular series of neutral signals, punctuated randomly by very occasional significant signals, to which the subject was expected to respond. In this the subjects sit looking at a black pointer 6 inches in length which normally moves like the second hand of a large clock in front of a white vertical surface.

Every second the pointer jerks on to a new position, one hundred of these movements making up the full circle. There are no scale markings or reference points of any kind on the background. The subject is told beforehand that every now and again at long and irregular intervals the pointer will move twice the usual distance, giving the effect of an unusually large jump forward. He is to watch for these movements, and press a response key as soon as he notices one.

Mackworth also used two other vigilance tests:-

(a) the synthetic radar test, where the subject had to watch for a small pin-point of light representing an "echo", which appeared occasionally at intervals which were arranged to be according to the same pattern of time as in the clock test.

(b) the listening test, where the subject had to pick out occasional longer sounds from a series of notes of the same intensity and pitch presented at regular intervals; again the temporal pattern was similar to that of the clock test.

From the clock test experiments Mackworth concluded that:-

(1) Substantially more signals are missed by the subjects after they have been working for half an hour at the task.

(2) This decline can be prevented by a rapid alternation of watchkeeping duties with some other work. Half an hour on and half an hour off is required, since one-hour spells are too long if this falling-off in performance is to be avoided.

(3) The deterioration in accuracy can also be prevented by supplying knowledge of results during the test spell.

(4) A sudden telephone message in the middle of the test spell produces a temporary improvement which lasts for only half an hour.

(5) Briefing the subjects beforehand to watch very carefully during a given period of the test does not affect the missed signal trend at all.

(6) On the other hand, 10 mg. "Benzedrine" (amphetamine sulphate) taken by mouth one hour before the start of the test successfully maintains the initial level of accuracy over the whole two hours of the test. This is a pharmacological effect and is not due to any possible suggestion effect of the knowledge that one has taken some tablets which might affect efficiency.

Mackworth also points out that "people differ greatly in their ability in the Clock Test, but this is not related either to visual acuity or to group intelligence test score. There is no known reason for these differences".

Mackworth's theoretical interpretation of this work is framed in terms of S-R theory. The double movement of the clock hand is regarded as the conditioned stimulus, the command 'now' during the practice period as the unconditioned stimulus, the pressing of the key as a conditioned voluntary response, while the knowledge of results given during the practice period is regarded as the reinforcing agent. The decrement in performance as the experiment proceeds is held to be due to partial experimental extinction, since the reinforcing stimulus is absent. The "principle of expectancy operating through self-instructions" keeps the performance up to a partial extinction level. Mackworth considers two other possible explanations - that the decrement may be due to secondary extinction, the extinction of the response to the short movements spreading over to the double movements; or to conditioned inhibition being established to the inter-stimulus interval. He rejects both of these alternatives, in view of his finding that knowledge of results seems to remove the vigilance decrement.

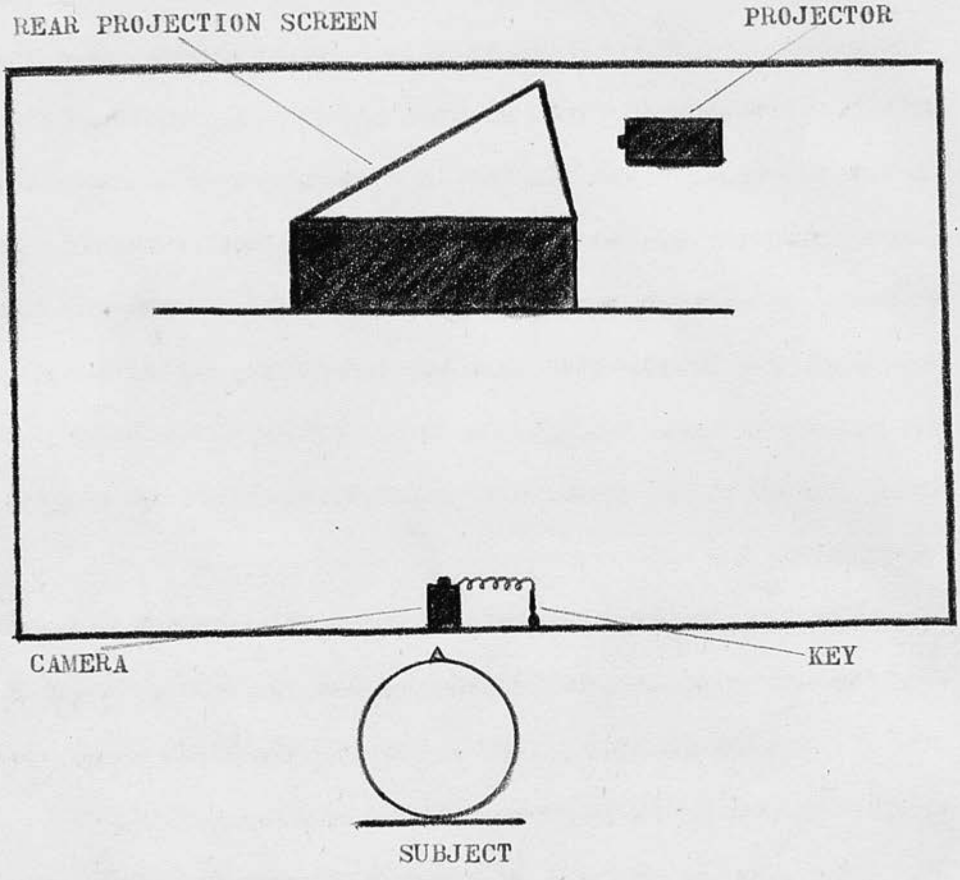
It will be seen that the tendency for a decrement to occur after some thirty minutes of continuous exposure to the vigilance situation appears to be a secure experimental findings. However, an experiment carried out by the writer (Fraser, 1950) failed to show this decrement after the first half-hour of working in a clock test situation where the subject was required to carry out the task for periods of one hour each under three conditions - when the clock face was vertical, at 45 degrees and horizontal, the operator's line of regard being normal to the display in each conditions. Consideration of this discrepancy led the writer to study the conditions under which the falling-off in performance tended to occur, and to advance a theory that deterioration in vigilance is a function of stimulus intensity, stimulus ratio and stimulus duration at the least. In the course of testing this theory the writer developed a form of vigilance task which avoided certain snags in the clock test apparatus, in that:

(a) The same pattern of signals was never repeated throughout the test period, thus avoiding any possibility of the subject learning the sequence.

(b) Since the presentation was entirely optical, there was no possibility of any auditory cue.

(c) The relation between the neutral and significant stimuli, was capable of variation in several dimensions.

In experiments, several variations of the one basic apparatus were used. This consisted of a simple projector system, prepared from a rear-projection vertical screen and an adapted single shot cine-camera. This projector was fitted with a motor which drove 16 mm. cine film continuously through a gate without a claw system, the drive being such that the film moved with a slightly irregular motion which was, however, perfectly reproducible.



VIGILANCE TEST APPARATUS - PLAN VIEW

Circular holes were cut in the film with a leather punch to form a moving display which was projected on to the screen placed between the subject and the projector. In the early experiments the holes cut in the film were of 2 mm. diameter, and there was one hole to each frame; at long intervals, randomly selected, a slightly larger diameter hole was cut. In all later experiments, there was one hole to each alternate or each third or each fourth frame. Recording was done by the subject pressing a button which operated the single-shot camera, and photographed the display. It should be noted that, as a fresh display can be used for each trial, there is no possibility of the subject learning the pattern of signals.

In the original vigilance test, the subject sat watching the display and pressed the button whenever he saw a large circle. For a ratio of circles close to unity, the test proved quite a useful variation on the traditional vigilance task.

Now, there is some evidence (Mackworth, 1950 ; Broadbent, 1951; Deese and Ormond, 1953) to suggest that the occurrence of a significant signal which is detected by the operator tends to reinstate the condition of alertness characteristic of the beginning of the watch period. It occurred to the writer, accordingly, that a vigilance task in which response to every significant signal was likely, but which involved some more subtle criterion of deterioration, would provide a more sensitive measure, and obviate the statistical, if nothing more, disadvantages of comparing performance of different operators in conditions where the stimulus situation is perhaps different because of this regenerative effect. Measurements of the discrepancies of the subject's judgments from the true centre, would, of course, bring in such things as reaction time differences and constant errors, so the measure chosen is the variance of the subject's estimates about his own mean estimate.

Preliminary research on this task (Fraser, 1951) showed that it formed a sensitive measure of vigilance, free from some of the drawbacks of other measures, and initial trials suggested that it might prove a useful index of stress.

APPROACH TO THE PROBLEM

The previous pages form the background to the present study, which is chiefly concerned with experimental work in the field to examine the changes in vigilance which occur under certain forms of stress. The work is arranged more or less in chronological order as the concepts developed, but for the sake of clarity the actual sequence of experimental work has been altered in places. The initial problem which presented itself to the writer may be framed simply - is it possible to demonstrate a significant change in performance of a vigilance task as the result of - or, more cautiously, - in association with a stress condition?

The writer has considered elsewhere (Fraser, 1951), (Fraser, 1954) the use of the term "stress" by various writers, and has suggested that it should be regarded as anything which in relatively extreme forms tends to disrupt normal functioning.

The essence of this approach is that in any experimental situation it is necessary to show that the "stress" conditions are liable to produce deterioration of performance; this deterioration need not necessarily occur on any given occasion, but there should be a significant probability of its occurrence.

It is improbable that many field studies are carried out, in Psychology at least, merely to test some theory devised as a result of laboratory investigation. In the writer's experience, most applied investigations are set up to answer some specific and immediate problem, such as the question of how long a radar operator may search for occasionally occurring signals before his efficiency in detection begins to fall off. When the investigator has answered his immediate problem, or as commonly happens, finds that he must ask himself a rather different set of questions about the subject under consideration, he frequently begins to examine the wider implications of his findings, and to see how far he can embody these findings in a comprehensive framework.

Then, if he has an inquiring mind and if his immediate employer is not breathing too heavily down his neck about the next problem but one on the experimental programme, he may have the opportunity to develop a systematic body of research. But, in general, there are two somewhat opposed aims in a given peice of applied research - that of the investigator, who wants to discover something of fundamental and general interest in the sphere in which he is working, and that of his sponsor, who wants the exact answer to some specific and limited question in the next week or so.

The present writer was fortunate in that, for a part at least of the time during which he was working on the present project, no precise answer was required of him to a specific question, and he was thus able to examine a fundamental problem in the way which fitted best his techniques and his interests. But, on the other hand, the mandate which he received to carry out research was not strengthened by any high-priority operational requirement, and he was therefore obliged to disturb as little as possible the normal working of the subjects in their environment, and to accept opportunities for investigation as they occurred. This meant that at times the experimental set-up fell short of the ideal of a closely controlled laboratory investigation. On the other hand, the facilities offered were considerably greater than those normally available to an investigator, on this side of the Atlantic, at least, and this is no small point in a country where the psychologist often has to work on a shoestring budget, with a grossly inadequate number of subjects and against a far too short deadline.

The simplest way of testing for the effect of a stress conditions on vigilance, assuming that we have a satisfactory test for vigilance, would seem to be:-

(a) to test the individual before applying the stress condition, and thus obtain a measure of his vigilance when he is presumably fresh.

(b) To put the subject through the stress condition.

(c) to test the subject immediately afterwards to observe the change in performance.

We take it as probable that a deterioration in vigilance is to be expected as a result of the stress to which the subject has been exposed. It is not axiomatic, since it may be that a certain degree of stress would actually improve performance, as a certain degree of tension is said to improve skilled performance. But if we put the individual through a long and rather wearing task we should expect that a sensitive measure of vigilance would show a deterioration rather than an improvement.

Using the vigilance measure in this way we are testing for the effects of stress by means of an interpolated test, and it has been argued by the writer (Fraser, 1953b) that "fatigue" effects do not show up well on interpolated tests. It is not necessary to go into the argument again here, except to say that the novelty value of performing the interpolated test tends to restore the level of performance to that of a fresh subject, for a time at least. Thus, it has frequently been found that interpolated tests often show an improvement after a presumably "fatiguing" activity.

We might, of course, expect an improvement in performance as the result of learning, and it is clearly essential to determine how great the learning effect will be for a given test by using a control group which does not undergo the stress condition, or by testing first immediately after the stress condition, and a second time hours or days later when the individual is fresh again.

That the interactions of this second arrangement can be complex is suggested by an experiment of Welford et al (1950). These investigators tested for fatigue in civilian aircrew, Radio Officers and Stewards of B.O.A.C., after flights ranging from 6 to 21 days. They conclude "that measurable impairment of performance is likely to occur at tasks met for the first time when fatigued, but that little or no impairment occurs when the tasks have been met previously in a non-fatigued state. However, as will be shown, no similar effect was detectable in the short-term flights examined here.

The precise effect of experience on the performance of a vigilance task is not clear. Mackworth (1950) suggests that there are no appreciable learning effects, while Broadbent (1951) suggests that certain individuals tend to show a marked improvement from the first to the second trial. Such individuals, he suggests, are more "fatigueable" than those who do not show this. The evidence for this he finds from a consideration of the tendency for such individuals to deteriorate during the task itself, and also from their life-histories, e.g. tendency to change jobs, not to persist in higher education when they have any choice in the matter, etc. Generalisation from a short laboratory task to a life-history is liable to be a hazardous occupation, but we may reserve judgment on this latter theory until more evidence is accumulated. The writer set out to test the relation between improvement from the first to the second trial and deterioration during the task itself using a group of naval ratings. Twelve subjects were tested under the same conditions as in Broadbent's experiments. When the average deterioration within runs was correlated with improvement between runs, a value for tau of 0.46 was obtained. $\frac{S - 1}{\text{S.E. of } S}$ was 1.98, so that this value is just significant at the 5% level.

Apart from this, the writer has not been able to observe any consistent effect of experience on performance of the task used throughout most of this study. In the later experiments, several control tests were given both before and after the stress condition, but in the earlier testing, where it was essential to avoid upsetting the normal routine of the subjects, the single control test was given sometimes before and sometimes after the stress condition. The next section, then, deals with the validation of the vigilance task in the field.

EXPERIMENT I - VALIDATION IN THE FIELD

INTRODUCTION

To validate the technique, the writer visited several airfields, some of which served piston-engined aircraft, and some of which served jet aircraft with a portable version of the equipment.

METHOD

The aim of the investigation was explained to the navigator subjects in a group before testing began, and an opportunity of asking questions was given. In particular, stress was laid on the fact that the experiment was intended to validate a potential research instrument, and was in no sense a psychiatric study. Subjects were then tested individually in a darkened room for a period of one hour on each trial. The procedure, as outlined in the preceding section, was explained to each subject, emphasising the importance of maintaining a consistent estimate throughout the test. The subject's watch was borrowed so that he would have no clear indication of the approaching end of the test (to minimise the possibility of an end-spurt). The subject was then left alone in the room for the duration of the test. After each trial he was asked for his comments.

The selection of operators was arranged by the squadron commander. This was not an entirely satisfactory feature of the experimental arrangements for two reasons. Firstly, since the normal flying schedule of the squadron could not be altered to any extent, it was not always possible to obtain subjects in the most suitable order for building up a neat experimental design. Secondly, there was sometimes reason to believe that subjects were being allocated to the experimenter because of the ease with which they could be detailed rather than in such a way as to give a truly representative sample of the squadron. (Thus in one station, all the subjects detailed were sergeants or flight-sergeants with no officers represented).

In the original experimental design it was intended to test each navigator three times - when fresh, after a short flight, and after prolonged flying. For the reasons given, it did not prove possible to randomise the order of serving under the three conditions in an entirely satisfactory way, but sufficient data were collected to permit of a comparison being made of performance when fresh with that after flying for a short period, and of the effect of giving the test first to fresh subjects with that of giving the test first to fatigued subjects.

RESULTS

(1) Comparison of performance of aircrew and naval ratings.

The first point which manifests itself clearly is the marked superiority of the aircrew. Although there is a wide distribution of variance scores on the test when fresh, the range from the best performance to the worst is 0.28 to 4.41 (in this method of scoring, a low score represents a good performance, and a high score a poor one). In a comparable group of naval ratings, the range of scores was from 0.5 to 19.2. Inspection of the individual results makes it clear that the difference between the two groups is considerable. If we consider, for example, the proportion of subjects in each group who achieve a variance score of less than 1.0, it will be seen at once from Table 1 that the performance of the aircrew is markedly superior. (A score of less than 1.0 may be regarded, rather arbitrarily, as a good score for a fresh subject).

Table 1

Comparison of Aircrew (Fresh) with Naval Ratings (Fresh).

Subjects	No. of Subjects	% age subjects scoring less than 1.0
Naval ratings	54	14.8%
Aircrew	36	41.7%

This result, of course, was not unexpected. The naval ratings used as subjects in the Psychological Laboratory probably represent a fairly random sample of the population, while flying personnel constitute a highly selected sample. But it is satisfactory that the test discriminated clearly between the two groups.

~~The distribution of scores for aircrew is shown in Fig. 1, and for the Naval Ratings in Fig. 2.~~

(2) Half-hour to half-hour trend in fresh subject.

One of the most interesting findings of Dr. Mackworth in the original vigilance experiments was that there is a significant increase in the number of missed signals in the second half-hour of such a test. This was confirmed strikingly by the writer (Fraser, 1951) using a form of the apparatus described in this study.

The same trend is seen even with this highly selected sample. Of 36 cases, 23 subjects showed a rise in the variance during the second half-hour of the test, and 13 subjects showed a fall. This gives a value for X^2 of 4.05, which with one degree of freedom is significant at slightly better than the 5% level (0.05 = 3.84, 0.02 = 5.41).

It has already been demonstrated (Fraser, 1952) that a reversal of this trend is usually found in subjects who show a poor overall performance, i.e. a high variance score. This is confirmed in the present experiment. Of those subjects who showed a reversal, 69% scored higher than the median value for the fresh group; and more than half (56%) of these scored more than 2, which, [redacted] is a very poor performance for this group. Indeed, the "tail" of the group is largely made up of subjects showing this reversal effect.

3. Effects of flying.

As already indicated, an interpolated test tends to show at a disadvantage when used as a test of fatigue, owing partly to the effect of novelty. It was stressed from the beginning that there should be no delay from the moment of touching down to starting on the test, but there is reason to believe that this ideal was not always obtained, particularly with the larger airfields. The effects obtained could, accordingly, almost certainly be improved by better organisation of subjects, and probably show the test at a disdavantage.

(a) Comparison of fresh subjects with those returning after a short sortie.

A short sortie is taken here to represent one or two one-hour spells on jet-engined aircraft during the day, or one similar spell at night, or a period of about four hours in a piston-engined aircraft. It is regretted that these rather diverse sorties have to be grouped together in this arbitrary way (particularly since an analysis of any difference between the conditions would be of so much interest), but this was necessary at this stage to obtain sufficient data to make an adequate analysis.

The rises and falls in the variance scores for these short sorties were comparatively small in almost all cases. Considering only the direction of the trend, for 26 cases, 15 rises were obtained after flying, and 11 falls. This gives a value for X^2 of 0.34, which with one degree of freedom falls considerably short of statistical significance. We must conclude that the test has not proved sufficiently sensitive so far to distinguish any effect resulting from a short flight. Whether this is due to the small number of cases, or whether there is in fact no appreciable impairment cannot be determined as yet. It is, however, worthwhile noting that almost all the navigators concerned maintained that they did not experience any feeling of "fatigue" after a short sortie.

(b) Comparison of fresh subjects with those returning after a long sortie.

A long sortie is taken here to represent a period of ten hours or more in a Lancaster or a Shackleton, or more than two one-hour spells on jet aircraft during the daytime, or more than one hour at night.

The results from this are unambiguous. Of a total of 30 cases, 24 of those returning from long sorties showed a rise in the variance score, to 6 showing a fall. With one degree of freedom this gives a value for X^2 of 9.63, which is significant at considerably better than the one per cent level (0.01 = 6.64, 0.001 = 10.83). In addition, of those showing a fall in the variance score, two subjects had been tested before flying, and this score was taken as their "fresh" score, but they had actually complained of having a bad night's sleep, which makes the criterion unsatisfactory. (It was not possible to obtain a second score for these subjects at the time). This almost certainly increases the validity of the test.

One individual case is worth recording at this point. The medical officer of one airfield referred for testing a sergeant who had been complaining of "fatigue" after flying. He was tested the morning after an eleven-hour flight (in a Shackleton) and returned an unusually high variance score, with reversal. Tested again next morning, after an intervening day of almost complete rest, he showed a greatly improved performance, with a score well above the median performance for aircrew, but again with reversal. This case is particularly interesting, since it seems to show three criteria of "fatiguability", viz. high variance, reversal and marked improvement between runs.

(c) Comparison of night flying and day flying.

When some of the early results were discussed, it was pointed out that the significant results obtained after night flying might be partly due to the fact that the subject was being tested at a time of day when he would show an impairment effect regardless of whether he had been flying or not. This argument, of course, does not affect the validity of the test as an instrument for measuring impairment, but it would undoubtedly be of interest if the two effects could be separated. One way to do this would be to test aircrew who had not been in the air at the same late hour as navigators who are tested after night flying. An indirect check on this, however, can be made by comparing the scores of aircrew tested after night flying with the scores obtained by navigators returning in the morning after a long flight (as happens in the case of the Coastal Command sorties in certain cases). The number of cases of the latter type recorded so far was quite small (five), but all these appeared to show a fatigue effect as great as that of the navigators tested at night. An adequate analysis of this point again awaits the accumulation of more data, but the evidence so far suggests that the effect of flying "fatigue" outweighs that of temporal position.

(d) Reversal effects after flyihg.

One effect of flying "fatigue" is interesting. It has already been mentioned that one of the best established trends in this kind of vigilance test is the rise in the variance score from the first to the second half-hour, and that a reversal of this trend often occurs in the case of subjects who show a poor overall performance. It was found, however, that with subjects taking the test after flying a long sortie, this trend was reversed, for good subjects as well as poor. Thus, for 30 navigators taking the test after a long sortie, a fall in the variance score from the first half hour to the second was observed in 21 cases. With one degree of freedom, this gives a value of 4.63 for X^2. This is just significant (0.05 = 3.84).

(5) The Welford-Brown-Gabb effect.

In 1950, as already indicated, Welford, Brown and Gabb carried out an experiment on fatigue in civilian aircrew, testing Radio Officers and Stewards of B.O.A.C. after flights ranging from 6 to 21 days. Their most interesting finding was that impairment of performance shown in subjects tested first after a trip appeared also when they were retested after at least 8 days' stand-down; and that, conversely, subjects initially tested after stand-down showed no impairment of performance when subsequently retested after a trip. From this they deduce that "measurable impairment of performance is likely to occur at tasks met for the first time when fatigued, but that little or no impairment occurs when the tasks have been met previously in a non-fatigued state".

No general effect of this kind was found in the present experiment. One night's sleep was enough to restore full performance on the task for the majority of the subjects, and there is no evidence that the order of taking makes any appreciable difference to the scoring. This is perhaps fortunate, since a carry-over effect of this kind would diminish the usefulness of the test as a measure of fatigue. It should be noted, however, that the periods of flying covered in Welford, Brown and Gabb's experiment were considerably longer than those of the present experiment, so that a cumulative fatigue effect might have been built up. It would clearly be of interest to test for a fatigue effect developing towards the end of a prolonged session of consecutive flights. There was, moreover, a slight indication of a carry-over effect in the more fatiguable subjects, i.e. those returning a high critical variance score. Again more data are necessary to decide whether this is a chance phenomenon.

It seems clear from these results that we can draw certain conclusions about the effects of flying on performance of the vigilance task under discussion.

(a) As far as piston-engined aircraft are concerned, no appreciable effects occur until after some seven to ten hours continuous flying, for the population sample studied.

(b) In the case of jet aircraft, it appears that three or more sorties during the day, or two or more sorties at night tend to produce a significant deterioration, in the sample studied. Since there was an interval of at least one hour between each sortie, we cannot compare this result directly with the finding above, since the effect may depend on, among other things, repeated take-off and landing (two of the most critical periods in flying), the usually rather boring waiting periods before and between flying, fitting and removing harness.

And, in addition, certain general conclusions about the performance of the test by aircrew:

(c) The tendency for a deterioration in performance to occur after half-an-hour on the task is observed even with this highly selected sample.

(d) The sample aircrew studied is distinctly superior in performance to a presumably more random sample of naval ratings.

DEFINITION AND MEASUREMENT OF FATIGUE

DEFINITION AND MEASUREMENT OF FATIGUE

The next stage of the investigation deals further with the problem of long-range flying in piston-engined aircraft. As already mentioned in the validation experiments, the technique was altered slightly in details as the investigation proceeded, so a simple better-or-worse criterion was adopted, using the chi-square test to determine significance. The next section provides more precise data on the degree of change involved. In general, the experiment is better controlled, but since it was originally devised for other purposes and the writer was offered the opporunity of taking part in it at a very late date, it is not as satisfactory as could have been desired. It should perhaps be pointed out here with an experiment of this nature involves a high degree of cost in money, time and training effort to the Command concerned, and is not easy to arrange for a single investigation.

Since the previous investigation the writer had considered further the problem of definition and measurement of fatigue. (In the original write-up of the investigation the non-committal term "deterioration of performance" was employed). The rest of this section discusses certain points in justification of the use of the term "fatigue" with a limited, objective meaning.

The work done in the field of fatigue is so considerable that it is not possible here to do more than list certain general findings. These are:

(a) There is no generally accepted definition of "fatigue". The term is used to cover such things as subjective feelings of weariness and boredom, physico-chemical changes in tissue, and objective deterioration in performance.

(b) Investigations on the subjective aspects of fatigue have not proved very fruitful, partly owing to very marked individual differences and to the difficulty of tying up these subjective reports with any objective measure. Bartley and Chute (1947), however, propose a definition of fatigue which is based largely on subjective factors. This is linked with a hypothesis that fatigue arises from conflict, the evidence for which is not overwhelming.

(c) Investigations of physico-chemical tissue changes have given clear-cut but limited results. It appears unlikely that the diminished irritability found in the physiologist's nerve-muscle preparations is directly related to the fatigue occurring in real life situations, except when the organism is approaching exhaustion.

(d) Most of the serious investigations of fatigue have been those concerned with objective changes in performance. These have been broadly of two kinds:

(i) Studies with tests given after "fatiguing" activities-interpolated tests.

(ii) Studies of deterioration during the task itself or during a similar synthetic task.

(e) Interpolated tests have shown, in a surprising number of cases and with unusual agreement between experimenters, little or no change after sleep deprivation, "mental work", muscular exercise, driving, flying and many other activities. The few exceptions have either been on tests which were directly related to the fatiguing activity, in which case they should perhaps come under the second category above, or have tended to show inconsistent results on repetition by other workers.

Occasionally, a significant effect has been demonstrated, without contradiction, but these cases, for reasons best known to the experimenters, do not seem to have been followed up.

(f) Studies of deterioration during performance of the task itself have proved more fruitful.

This position is far from satisfactory. The failure to achieve a definition of fatigue which will have operational validity, means that a serious gap exists in the treatment of skilled performance and, indeed, in all those cases where fatigue is considered to be operative.

Moreover, in spite of the difficulty of experimental demonstration, there is some indirect evidence that deterioration in performance as a result of long periods of skilled activity, or, more cautiously, in association with long periods of skilled activity, is responsible for many accidents. It seems well worth while to examine the question again to see if it is possible to explain the comparative failure of laboratory investigations and to achieve an operational definition.

It is clear that the devising of a good, objective test for "fatigue" will present a number of difficulties, not the least of which is the well-established fact that the novelty value of an interpolated test tends to mask any deterioration. Again, as we have seen, none of the simple physiological tests appear to be of great value, since they are effective only when the subject is nearing a state of exhaustion, and, as Bartlett (1948) has pointed out so clearly, the raggedness of timing and increased tolerance of error which characterise the early stages of breakdown in skilled performance occur long before the stage of exhaustion.

We may accept as a tentative definition that "fatigue" refers to a deterioration in performance occurring under stress conditions where one of the most important factors is the length of time for which the subject has been engaged in continuous performance. The writer's thesis, is then, that the characteristics of "fatigue" can be derived by considering it as a special form of stress condition and applying the general principles established in that field. The most relevant of these are (Fraser, 1951, 1954, 1956a):-

(a) The effect of increasing stress beyond a certain level is to produce increased deterioration in performance, the falling-off being roughly proportional to the increase of stress. A simple mathematical relationship can be found between the degree of stress and the falling-off in performance in several types of stress conditions (Mackworth, 1950; Fraser, 1951; Conrad, 1951).

(b) The effect of stress on performance is a differential one, the performance of the less skilled subjects deteriorating much more rapidly than that of the more skilled subjects. (In other words, if we plot impairment of performance against degree of stress, the curves of the less skilled subjects slope earlier and more steeply. This, of course, implies some criterion to indicate a subject's degree of skill; this has to be worked out for any experimental situation.

(c) A general effect of stress beyond a certain level is to produce increased variability of performance. The less skilled subjects again tend to show this increased variability more.

(d) The effect of a stress condition is to a considerable degree dependent on the knowledge of results or knowledge of performance available to the operator. In general, the more complete, the more precise, the more immediate the knowledge of results the less the performance is adversely affected by a stress condition.

Applying this theoretical framework to the problem of fatigue, we can postulate that an adequate test for fatigue should possess the following properties:-

(a) In view of the fruitless results obtained from simple measures of isolated function the test itself should involve continuous serial performance.

(b) The conditions of application of the test should be such as to render it a stress condition, e.g. it should be applied at such a speed, with such a load or for such a duration that we would expect a deterioration to appear on the task itself.

(c) In view of the clear evidence that knowledge of results tends to stabilize performance under stress, the test should involve a minimum of such knowledge.

If the foregoing analysis is substantially correct we have at least two reasons why practically all interpolated tests for fatigue fail to give significant results. In the first place, since they seldom involve a real stress condition, and since the novelty value of an interpolated task is high, the subject is not really extended. In the second place, they usually provide the subject, by interwal "feedback", with sufficient knowledge of results to maintain adequate performance.

Fatigue is far too useful a concept to be either banished completely from all precise discussion, as Muscio suggested, or used to cover a variety of subjective states of the individual, as Bartley and Chute infer. Since the subjective aspects of fatiguing situations are already very adequately provided for by such terms as weariness and boredom, we ought to differentiate the term completely from such subjective factors, and give it an objective, if at first limited definition.

The following objective definition is, therefore, proposed!

Fatigue - A disorganisation of receptor-effector co-ordination, resulting from exposure to high speeds, loads, duration, anxiety, conflict, sleep deprivation or other form of stress conditions. Manifested most clearly in operations where knowledge of results is minimal and duration of task is prolonged, so that the normal human ability to compensate is inadequate to maintain co-ordination.

From this point on, the variability of judgments measured under vigilance task conditions is described as "z-function".

EXPERIMENT 2 - COMPARISON OF DAY AND NIGHT FLYING

PROCEDURE

The writer was given the opportunity of taking part in this research at the last moment, so that it was not possible to give all the subjects the usual initial "buffer" trial to offset the atypical scores made by some subjects in the first trial of the series (Broadbent, 1951, Fraser, 1956).

The flight trials of this stage of the research comprised originally sixteen 15-hour sorties, each with a different crew, and each including substantially similar operational requirements. Half of these were by day, and half by night. The whole series was completed in ten days, but three further flights were added at a later date. One member of each crew, either a pilot or a navigator, was tested by the writer immediately after landing. Each subject was also given a control test; precautions were taken to ensure that each subject had had a good night's rest and was reasonably fresh when he undertook this control test.

Certain changes were made in the administration and scoring of the task since the previously reported investigation with the object of increasing its sensitivity to fatigue. The most important of these was to present ten large circles during the hour test instead of twenty. This was done because there is good reason to believe that the occurrence of a significant signal which is observed tends to reinstate beginning-of-watch performance (Mackworth, 1950; Broadbent, 1951; Deese and Ormond, 1953) - an important point with the present technique where it is rare for a subject to miss a signal. This change makes the calculations of the variance for each half-hour of the test rather pointless, since a variance score computed from five readings would be unreliable as a sample estimate. It also makes a study of the reversal phenomenon impracticable. However, it was felt that the possibility of an increased vigilance deterioration would compensate for these points.

Although the experimental work on z-function in this research was carried out solely by the writer, a series of other measurements were made by physiologists and other medical research workers on other members of the crew taking part in these flights

Answers to the following questions were sought:

(a) Does the technique again show a significant rise in variance scores after flying compared with the scores obtained when fresh?

(b) If so, is there a significant difference between changes occurring after night flying as compared with day flying?

(c) Is there any evidence to settle the point raised earlier that the significant results obtained after night flying may be partly due to the fact that the subject is being tested at a time of day when he would show a fatigue effect regardless of whether he had been flying or not?

(d) Do the results support the general thesis that increasing stress tends to produce increasing variability of judgment both within and between subjects?

RESULTS

(1) Table I shows the variance scores obtained before and after flight by all 19 subjects. It will be seen that there is a significant difference ($P = 0.03$) between the control and experimental scores for the day-flying group and a more significant difference ($P = 0.007$) between the control and experimental scores for the night-flying group.

(2) Comparing the rise in variance score for the night-flying group with the corresponding rise for the day-flying group, we find a significant difference ($P = 0.01$) between the two groups.

(3) Since the subjects of the day-flying group were tested at midnight, and the subjects of the night-flying group were tested about 8 a.m., this result indicates that the significant rise in z-function obtained after short night flights in the earlier experiments [redacted] was not due to the fact that the subject was tested at a time of day when he would show a fatigue effect in any case.

(4) Comparing the variance of the "day" control scores with the variance of the "day" experimental scores we obtain an F-ratio which is significant at the 2% level. Comparing similarly the variance of the "night" control and experimental scores, we obtain an F-ratio which is significant at better than the 0.1% level.

TABLE I

Control and Experimental Scores obtained by 19 subjects, together with significance of the difference between means using t-test technique.

Subject	Control Scores		Experimental Scores	
	Day (D)	Night (N)	D	N
1	1.46		1.13	
2	0.28		0.81	
3	0.58		1.05	
4	1.13		1.05	
5	0.50		0.96	
6	1.03		0.83	
7	1.28		4.01	
8	1.57		1.56	
9	0.21		1.90	
10	1.26		2.40	
11	0.70		1.03	
12		1.57		8.63
13		0.68		1.84
14		1.36		3.27
15		1.13		2.97
16		1.60		2.27
17		0.72		7.03
18		2.11		2.52
19		2.65		8.41
Mean	0.91	1.48	1.52	4.62
S.D.	0.48	0.67	0.96	2.89

t_{10} (C-d)	2.20	[(p = 0.03 (approx.)]
t_7 (C-N)	3.23	[(p = 0.007(")]
t_{17} (C-D, C-N)	2.50	[)p = 0.01 (")]

If we take it therefore, that night flying is more stressful than day flying of equivalent length, the results fit neatly enough into the general thesis that increasing stress tends to produce increasing variability of timing, both within and between subjects.

DISCUSSION

The result agrees well with the introspective observations of the aircrew; feelings of weariness and strain were reported much more frequently from the subjects on night flying. Some comments made by Squadron Leader G. Melvill Jones, of the R.A.F. Institute of Aviation Medicine, are relevant here:

"During the night flights a growing desire for sleep was found among crew members, being most marked in the hours just before dawn. A strong effort of will was necessary to maintain constant attention applied to the flying task. Although a safe standard of flying could be maintained, simple observation of flight instruments was enough to show that accuracy in flying a steady course deteriorates towards the end of a 12-15 hour flight".

It seems safe to conclude that one factor in the greater fatigue effect manifested after night flying is the disturbance of the normal diurnal rhythm. Although aircrew are supposed to rest in the morning before a flight, there is good reason to believe that this ideal is not easy to achieve. As Melvill Jones remarks:-

"For a number of reasons it is impossible on a normal station to sleep through the morning before an exercise ..... In practice ..... the morning rest becomes little more than a 'lie-in', sooner or later terminated by boredom".

The experiment also settles the point raised earlier - that the significant results obtained after night flying might be partly due to the fact that the subject was being tested at a time of day when he would show a fatigue effect regardless of whether he had been flying or not. In fact, considering that the subjects in the day-flying group had done 15 hours flying and were being tested from midnight to 1 a.m. it is remarkable that the fatigue effect after the flights is so small.

EXPERIMENT 3 - CUMULATIVE FATIGUE AND VIGILANCE

INTRODUCTION

As indicated by the earlier work, the long-range flights of Coastal Command tend to show a fatigue effect commencing some 7 to 10 hours after take-off; and the previous section indicates that there is a significantly greater effect after night flying than after day flying of equivalent length. It is not easy to obtain a significant effect in these Coastal Command flights with a small number of subjects, since some of them seem to show little or no fatigue effects. This may be due, at least in part, to differences in flying conditions and to the fact that it is possible to obtain rest and even sleep during flight. Moreover, as in all such field investigations, it is assumed that the control scores are taken from fresh subjects, but of course it is impossible to exclude fatiguing factors extraneous to flying.

PURPOSE OF THE INVESTIGATION

The present investigation was undertaken to study the effects, if any, of cumulative fatigue as the result of flying four 15-hour sorties by night with one day's rest between each. Since the investigation was highly organised to test a number of other variables, and it was necessary to accept certain limitations as far as the vigilance tests were concerned. In particular:-

(a) Only 40 minutes could be devoted to each test period, instead of the customary one hour.

(b) Two subjects had to be tested in succession after each flight. This is not ideal, since there is some evidence that the maximum fatigue effect tends to pass off about an hour after landing.

(c) Only a small number of subjects could be tested.

PROCEDURE

Six subjects were given three control tests at Farnborough some days before the experiment proper, which was conducted at Coastal Command base.

Each subject was then tested either immediately after each flight or 45 minutes later, according to whether he was first or second for testing. Each of a given pair of subjects from one particular crew took the test first or second alternately after each flight. Each subject was asked to give his subjective impression of how he felt both before and after each test, and he was also asked to comment on the severity of the trip.

RESULTS

Table 1 shows the control and experimental scores for the six subjects, together with the mean scores and difference between means.

TABLE 1

Z-scores of subjects flying repeated sorties

Subject	Control	Sortie 1	Sortie 2	Sortie 3	Sortie 4
1	0.93	2.0	1.7	24.9	4.0
2	1.83	3.4	2.6	3.2	4.7
3	2.72	1.8	4.2	2.9	3.8
4	1.27	3.64	3.49	4.59	4.20
5	1.19	1.61	0.93	2.45	1.35
6	0.53	1.44	6.43	4.07	7.20
Mean	1.41	2.31	3.22	7.02	4.21

Difference between Control and All Days Combined $0.001 < p < 0.01$

" " Control " Day 4 $0.001 < p < 0.01$

" " Day 1 " Day 4 $0.01 < p < 0.05$

(a) There is a highly significant difference between the mean control score and the mean of the combined scores for all the experimental days.

(b) Inspection of the mean scores suggests that there is a progressive rise in z-function from Control to Trip 1, from Trip 1 to Trip 2, and from Trip 2 to Trip 3, with a slight fall in z-function from Trip 3 to Trip 4. Since, however, the only significant differences are between Control and Day 4, and between Day 1 and Day 4, this cannot be regarded as proved.

Inspection of the raw data suggests that the peak score on Day 3 is considerably influenced by the very high score of Subject 1. If we omit this subject's scores we find that there are significant differences between Control and Day 3 also. (The legitimacy of this omission is confirmed by the fact that this subject was later found to have been unwell on the third trip).

(c) The subjective reports of the subjects were grouped into six classes ranging from:- (0) No different from fresh control, to

(+++++) Extreme degree of fatigue.

The z-scores of the subjects were grouped into the appropriate classes and the correlation between them was calculated. This correlation proved to be highly significant ($p < 0.001$).

DISCUSSION

It seems likely that a larger number of subjects and a more generous allocation of the test time would have yielded still more information. Even on the present evidence, however, it seems reasonable to conclude that :-

(a) Post-flight vigilance after the fourth sortie is significantly lower than after the first sortie. There appears to be some tendency for a progressive decrease in vigilance to occur as we go from the first to the fourth sortie, but this is not demonstrated as significant in the present experiment.

(b) Flying four 15-hour sorties at night with a day's rest between each sortie is likely to produce considerable fatigue in the crews concerned, manifested both as objective deterioration in post-flight vigilance and as subjective reported weariness and exhaustion.

(c) Subjective fatigue and objective deterioration in vigilance are significantly correlated in this investigation. It would be oversimplifying the position to identify the two, however, since :-

(i) Several subjects reported that the final trip was markedly less fatiguing than the third trip, but this was not reflected in their z-scores.

(ii) The earlier work has shown that a significant deterioration in vigilance may be unaccompanied by subjective reports of fatigue.

RELATION OF VIGILANCE TO OTHER MEASURES OF PERFORMANCE

It is clear from the previous sections that the test demonstrates increasing variability of judgment in association with various apparently fatiguing conditions. It would, of course, be very interesting to know how this change in z-function correlates with other measures, performance or physiological.

The previous investigation provides some information on this point. The experiment on cumulative fatigue was organised to test a number of other variables besides vigilance. Some of these were:-

(a) Continuous recording of altitude, direction and turbulence during certain sample periods of each pilot's shaft, strictly according to a pre-arranged schedule.

(b) Two simulated submarine searches of predetermined pattern, the pattern being unknown to those on board.

(c) One daylight pilot bombing episode in each flight plan.

(d) A number of physiological tests considered to bear some relation to stress.

If the hypotheses discussed are valid, it should be possible to predict those of the above measures which are most likely to show the effects of fatigue. Grading them in order of knowledge of results, we can put pilot bombing first, since the pilot is given immediate and quantitative knowledge of the relative success of his performance after each pair of bombs has been dropped. Straight and level flying may be rated next, since the pilot has immediate information about his present altitude and direction, but does not have this continuously presented. We may perhaps place submarine search third in this list.

Finally, the z-function measure, in view of its almost total lack of knowledge of results, may perhaps be rated as the most likely to show a marked fatigue effect; against this must be placed the fact that it has to be given as an interpolated test.

In addition to knowledge of results, we must also consider motivation. From this viewpoint, pilot bombing again ranks high, in view of its closeness to the real-life situation, and the fact that it occurs as something of a highlight in an otherwise monotonous routine. Submarine search probably has also a fairly high motivational value from the same point of view. Pilot motivation, according to certain experienced pilots, is minimal during straight and level flying.

In view of these variables in the situation, it is difficult to predict exactly the relative degree of fatigue which may be expected, but the above analysis provides us perhaps with a reasonable guess. We might expect that fatigue effects would be easiest to demonstrate in the z-function task, and hardest to demonstrate in the pilot bombing, with the submarine search and the straight and level flying coming between the two.

Finally, we might reasonably expect a deterioration to occur in performance of any of these measures without a demonstrable change in the physiological measures.

It is not possible to give full details at present of the actual results obtained in this investigation, but certain broad trends were clearly evident:-

(a) As already indicated, the z-function measure showed a significant deterioration in vigilance as a result of the four flights, and also showed a significant deterioration between the beginning and the end of the experimental period.

(b) Submarine search showed no significant deterioration between the beginning and the end of the experimental period.

(c) Straight and level flying showed a significant falling-off between the early and late period in each flight, but no falling-off between the beginning and end of the experimental period. There was, however, a significant increase in variability of performance between the early period and the late period in one flight, and a significant increase in variability of performance when the first two flights were compared with the second two flight.

These results, at least, do not contradict the analysis of Section 5, and suggest that further investigation along these lines would be fruitful.

One point comes out clearly from this comparison. A fair number of physiological measures which were considered likely to indicate the effects of exposure to stress were applied to the subjects of this investigation - a complete haematological analysis and urinary analysis. The only significant effects detected were a eosinopenia when all flying days were compared with intermediate resting days, and an increased excretion of uropeosinogen when all flying days were compared with resting days. It is apparent that we can have marked changes in performance and subjective states of subjects without any very extensive physiological changes.

COMPARISON OF PISTON-ENGINED AND JET AIRCRAFT

The initial experiment in this study showed that a significant deterioration in vigilance was detectable after 3 one-hour sorties during the day or 2 one-hour sorties at night. It was pointed out that the fatigue effect might be partly dependent on repeated take-off and landing, waiting periods before and between flying, fitting and removing harness.

One finding here is of particular interest. Many of the subjects in the jet crews maintained that they did not feel particularly tired after 2 or 3 one-hour sorties. This contrasts with the general tendency for subjective feelings of weariness and boredom to become generally evident after 8-9 hours flying in piston-engined aircraft. Although our definition of fatigue is framed objectively, the correlation, or lack of correlation, between the objective measure and the subjective experience is of great fundamental interest. It has been pointed out by Bartlett (1948) and the writer (Fraser, 1953) that a critical stage in fatigue may be the period during which the subject's objective performance is deteriorating while subjectively he believes that he is doing as well as or better than ever.

While the pilot of the piston-engined aircraft has a relatively leisurely task, the position of the jet fighter pilot is different.

The high speed of the jet, the long run required for take-off and landing and the limited fuel all combine to make his one-hour sortie exacting. When we consider also that the jet planes are reported as being easy and light to control and relatively free from noise as far as the crew is concerned it seems likely that the fatigue induced will be due to tension and anxiety, rather than to any physical strain involved. It is not suggested here that the old distinction between "mental" and "physical" fatigue should be revived. As far as this study is concerned, the term fatigue is reserved for objectively measurable deterioration regardless of the precipitating factors. But it is worth while considering how far we can separate different patterns of fatiguing activity.

On the one hand, then, we have a fatigue effect developing after a continuous flight of ten hours or more in a large, heavy noisy aircraft involving a fair degree of physical effort to handle, and accompanied by marked subjective weariness in the aircrew. On the other hand, we have a fatigue effect developing after 2 to 3 one-hour sorties in a small, smooth-flying, quiet aircraft involving little effort to handle, and unaccompanied in many cases by subjective reports of weariness.

The next investigation was therefore one to examine further whether a significant deterioration in vigilance occurs in jet flying over a period of 3-5 hours continuous jet flying.

EXPERIMENT 4 - COMPARISON OF CONTINUOUS AND INTERMITTENT JET FLYING

PROCEDURE

The aim of this investigation was to check how far it is possible to show a significant decrease in vigilance in aircrew flying (i) a continuous sortie of 3-4 hours in jet aircraft

(ii) 3 one-hour sorties in jet aircraft.

(i) For the first part of the investigation, aircrew were tested before and after flying for a 3-4 hour continuous sortie in a jet bomber carrying out a routine exercise during this period.

RESULTS

This part of the investigation was carried out over a fairly lengthy period during which there were changes in both the apparatus and the conditions of administration. It is only possible therefore to assess individual scores on a better - worse criterion. Out of a total of 26 subjects tested over this period, 16 subjects showed a rise in z-function in comparison with their control scores, and 10 showed a fall in z-function. This gives a value for chi-square of 0.96 which falls considerably short of statistical significance. We must conclude that the evidence does not suggest that 3-4 hour continuous flight in a jet bomber of this kind is liable to produce fatigue in the aircrew.

Few of the changes were at all marked, and almost all members of aircrew concerned maintained that they had no subjective feelings of tiredness.

(ii) For the second part of the investigation, a small group of pilots flying jet-engined fighters were tested, under conditions which were carefully controlled, some time before and immediately after 3 one-hour sorties (all during daylight).

RESULTS

The individual scores are given in Table 1.

It will be seen that :-

(a) No subject has a control score of greater than 1.0.

(b) All subjects show a rise in z-function after flying.

(c) In spite of the small number of subjects, the mean rise in z-function after flying is significant at the 2% level .

TABLE 1

Comparison of Control and Experimental Scores for Subjects Flying for 3 one-hour Sorties in Jet Fighters

Subject	Control	Experimental
1	0.85	1.2
2	0.95	1.23
3	0.81	1.35
4	0.9	1.87
5	0.57	0.67
6	0.81	1.64
Mean	0.82	1.33

Difference between Control and Experimental Significant at 2% level.

DISCUSSION

The data do not suggest that jet flying per se is more likely to induce fatigue than flying in piston-engined aircraft. Unfortunately we cannot from the present evidence say conclusively that the fatigue effect obtained in the case of the jet fighter subjects is due to the fact that they have to do three separate sorties with the consequent delays and multiplication of critical periods.

It may be that the jet bomber studied in the first part is inherently less likely to induce fatigue in the crew than the smaller jet fighter. (In fact the performance of this aircraft was highly praised by almost all the crew members).

Resolution of the part played by the multiplication of sorties awaits the development of longer range fuel tanks for the jet fighters studied above. It should then be possible to make a careful study of how far the fatigue effect is dependent on this multiplication by testing pilots after three-hour continuous sorties and after 3 one-hour sorties in the same type of plane .

HEAT STRESS AND VIGILANCE

HEAT STRESS AND VIGILANCE

INTRODUCTION

The preceding sections indicate the utility of testing for the effects of prolonged flying by means of this z-function measure given under standard vigilance task conditions. However, the length of the test makes it rather difficult to use in many field situations. A shorter test would not only be easier to fit into a compressed timetable, but would perhaps be less likely to impair motivation in the subject.

The simplest way to achieve this would obviously be to speed up the display so that the subject has the same number of randomly presented significant signals in, say, fifteen minutes.

This was actually tried in two classes of fatigue situations, but no significant results were obtained; the experiments, accordingly will not be described further. It appears that some other method of increasing the stress on the subject is necessary to demonstrate impairment as the result of a stress situation.

In view of the presumably marginal nature of the impairment, if any, involved, it seemed likely that speed stress might prove the most effective additional factor in this case. Moreover, it was felt desirable to re-introduce a task close to the original vigilance conditions. It has already been pointed out that the clock test involves certain disadvantages in respect of repetition of signal patterns, and the possibility that some auditory cue may be provided by the apparatus. The test used (and called the Cambridge visual vigilance test) in the present investigation was developed by the writer to provide a comparable vigilance task to the clock test, but as already mentioned in the case of the original vigilance task, with the following advantages:-

(a) The same pattern of signals was never repeated during the test period, thus avoiding any possibility of the subject learning the sequence. (However, unlikely it may seem to anyone who has ever sat through the clock test that the pattern of signals is capable of being learned, this did occur with one or two of Pepler's subjects in Singapore).

(b) Since the presentation is entirely optical there was no possibility of any auditory cue.

(c) The relation between the neutral and significant stimuli was capable of variation in several dimensions.

APPARATUS

The signals consisted of circles of light which appeared on a screen regularly every five seconds. At irregular intervals, randomly chosen, a larger circle appeared. The subject's task was to watch out for these larger circles, and to record them photographically by pressing a button. The ratio between the circles was found in earlier work to be an important factor in the 'fatigue' effect (Fraser, 1951); in the present experiment it was 3:2 (diameter).

In the modified version of the apparatus used for these experiments the signals occur every two seconds and remain on the screen for approximately 400 milliseconds. The test lasts only fifteen minutes (during which ten significant signals are presented), as compared with the customary one to two hours of the orthodox vigilance task, but it was hoped that the speed of presentation and the short duration of the signal would impose sufficient stress on the observer to measure impairment of vigilance under stress, even in this short time. There are three possible criteria of performance - a) Fast reactions, which implies the recording of the significant signal while it is still on the screen; b) Missed signals; c) Errors (button pressed for a neutral signal in mistake for a significant one).

EXPERIMENT 5 - DONCASTER RESCUE TRIALS

BACKGROUND TO THE INVESTIGATION

The Doncaster Rescue Trials (Lind et al., 1955) were primarily designed to study physiological changes occurring during exposure to heat under conditions resembling those of rescue work during mine accidents. The experimental conditions comprised ambient temperatures ranging from 85°F to 100°F (saturated) viz. - 85°F, 87.5°F, 90°F, 92.5°F, 95°F, and 100°F effective air movement - 150-175 ft./min.; period of exposure - 65 mins., or until rectal temperature reached 101.8°F (approx.), or until the subject elected to leave; work - one 30 mins. period consisting of five 4-mins. periods of walking at 2 m.p.h. carrying a 50 lb. sandbag and apparatus weighing approx. 40 lb., five 1 min. rest periods and one 5-min. period walking at 2 m.p.h. carrying a load as above, a 5-min. rest period, and another 30 min. period of heavy work, which consisted of constructing a sandbag stopping of 24 sandbags, each 50 lb., carrying each bag singly for 15 ft. to the construction point. The actual energy costs were:

(a) for the first 30 minutes 171 Kcal/m^2/hr.

(b) for five minutes rest 55 " " "

(c) the second 30 minutes 248 " " "

One special feature of the investigation was that subjects were exposed to the experimental condition for a period of 65 minutes, or, as happened at the higher temperatures, until they reached a physiological end-point. Although this was necessary to simulate the real-life situation, it complicates the comparison of earlier and later experimental conditions.

The following hypotheses were specifically tested during the investigation:

(a) That exposure to heat stress produces a significant deterioration in performance of the Cambridge visual vigilance task given under speed stress.

(b) That increasing ambient temperature produces increasing deterioration in the task.

PROCEDURE

Each subject was tested twice, before going into the chamber (control) and ten minutes after he came out (experimental) on each occasion. Each subject was given a very short trial run before each testing, whether control or experimental. After each run he was asked for his comments. In general the subjects appeared to be co-operating well, and their comments indicated that they found the test interesting and considered it likely to provide realistic information. This is particularly interesting since these are not the most usual comments of a subject who has just done an orthodox vigilance task.

RESULTS

(a) General Effect on Performance

As already indicated, a preliminary examination was made of several factors involved in the administration of the technique, and certain changes were made as the result of experience. This fact, of course, makes an absolute comparison of the scores obtained in the early and later parts of the investigation rather difficult as far as the delay in reacting is concerned, although there is no difficulty in comparing the number of missed signals. We can, however, examine the direction of trend of the experimental scores as compared with the control scores throughout, including all the criteria, to test the null hypothesis that no significant trend is involved. Excluding all cases where there is no change and omitting 14 cases where there was some ambiguity, this procedure gives a value for chi-square of 7.22 (corrected for continuity), which for one degree of freedom is significant at better than the 1% level (0.01 = 6.64, 0.001 = 10.83).

In accordance with the definition recommended earlier, it seems clear that we can now use the term "stress" legitimately to denote a range of ambient temperature conditions around the values studied in the investigation. The design of the experiment does not permit a rigorous examination of the degree of the heat stress conforms to the basic stress paradigm (Fraser, 1951, 1956), but further analysis suggests certain interesting possibilities for future investigation.

(b) <u>Missed Signals</u>

A missed signal in this task, represents a serious failure of vigilance comparable, perhaps in a real-life situation to the failure to observe an important and clearly discriminable signal for which the observer is waiting. Thus, although the total number of missed signals in these experiments is small, the fact that they occur at all is important.

Table 1 shows the occurrence of missed signals under five conditions. The control score represents the results of the tests which were given to the men when fresh; no signals were ever missed by any of the subjects during the control experiments.

Comparing the control and total experimental conditions, it will be seen that there is a highly significant difference between the two. Applying the usual significance test for a Poisson distribution which is appropriate for this class of data, it will be seen that there is less than one chance in ten thousand that the results could have occurred by accident.

Considering now the scores obtained at different temperature, the table shows that there is a highly significant difference between the scores obtained under Group 2 conditions and those of Group 1, Group 3 and Group 4.

There are no significant differences between the scores obtained under Group 1 conditions, Group 3 and Group 4 conditions. There is a significant difference between the scores obtained under the Group 1 condition and the control, and a highly significant difference between Group 2 and the control, but no significant differences between control and Groups 3 and 4.

Group 1 represents the pooled scores for the 85° and 87.5° temperature; Group 2 the pooled scores for the 90° and 92.5° temperatures, and Group 3, the pooled scores for the 95° and 100° (light work) temperatures; Group 4 represents the 100° temperature with heavy physical work. It was necessary to pool the temperatures in this way to obtain a sufficiently large number of cases in each group. Only twelve cases were available for the 100° (heavy work) temperature.

Taking the missed signals as sole evidence of a deterioration in vigilance, it appears that vigilance is significantly reduced during the Group 1 conditions, (as compared with the control level) more so in the Group 2 condition, improves almost to the control level with the Group 3 condition, and is equal to the control in the Group 4 condition - the severest of all. At first sight these results seem to contradict the hypothesis of diminishing vigilance with increasing ambient temperature. If we consider the circumstances of the experiment, however, the result is less paradoxical. As the temperature increases, the subject is withdrawn from the heat chamber after a progressively shorter period of exposure. If we assume that deterioration in vigilance is a function of a) temperature, b) duration of exposure, and c) work done, it is apparent that at the lower temperatures the subject is staying in the chamber for practically the whole 65 minutes, but at the high temperatures he is being withdrawn after anything from 27 to 45 minutes.

Moreover, since the heaviest physical work occurs during the last half-hour of the exposure under Groups 1, 2 and 3 conditions a subject withdrawn after less than 35 minutes will do no hard physical work at all.

The assumption that duration of exposure is a factor in producing deterioration of vigilance may seem unnecessary, since the fact that progressively less physical work is included at the high temperatures might be enough to account for the improvement in vigilance by itself. It was to clarify this point that the final Group 4 condition was added to the experiment. In this condition the subject is given heavy work from the beginning of the exposure to the 100° temperature. The results show no deterioration in vigilance in the form of missed signals, although it was unfortunately not possible to obtain more than 12 subjects at this final stage.

It might, however, be objected that the results do not show that physical work affects vigilance at all, since the scores on the missed signals are approximately the same for both the 100° (light work) and 100° (heavy work) conditions. But as will be seen in the next section, the 100° (heavy work) condition does produce a marked decrease in the number of fast reactions.

TABLE 1

Missed Signals and Ambient Temperatures

	No. of Subjects	Missed Signals
Control	72	0
Group 1 (85 - 87.5	20	3
Group 2 (90 - 92.5)	20	9
Group 3 (95 - 100 (L))	20	2
Group 4 (100 (H))	12	0

TABLE 2

Comparison of Temperature Groups for Significance

Group compared			P
Control	v	Gps. 1-4	0.0001***
Control	v	1	0.04*
"	v	2	Almost zero***
"	v	3	0.08 N.S.
"	v	4	-
Group 2	v	1	0.04*
" 2	v	3	0.014*
" 2	v	4	0.017*
" 1	v	3	-
" 1	v	3	0.11 N.S.
" 3	v	4	0.16 N.S.

* - Significant

*** - Very highly significant

N.S. - Not significant.

Fast Reactions

No very noticeable change occurs in the properties of fast reactions until the 100° (heavy work) condition is reached. At this point there is a quite dramatic drop. For this condition, as noted, only 12 subjects were available and although 9 of these* showed a drop in the number of fast reactions this is not quite significant ($X^2 = 3.25$; $p = 0.075$). Only the better-worse criterion can be used here in view of minor changes in the technique which make it difficult to average all scores for this criterion.

DISCUSSION

The results from this experiment suggest that it is possible here again for an interpolated test to show a significant effect as the result of exposure to another kind of stress. In this case, as in the previous cases, we are testing during the recovery period. It seems not unreasonable to suppose that testing during the exposure period would give a more striking result. And, clearly, the effect is obviously due, as far as the missed signals are concerned, to a small number of subjects. Before examining the problem further, it seemed desirable to check that the effects could be obtained again under perhaps rather different circumstances. To do this another investigation was made, this time at the Climatic Laboratory of the Institute of Aviation Medicine, using aircrew as subjects and with more flexible experimental arrangements.

* One subject showing no change.

EXPERIMENT 6 - FARNBOROUGH INVESTIGATION OF HEAT STRESS

HYPOTHESES TESTED

Examination of the data in the previous section shows that, out of 720 significant stimuli presented under the control conditions no signals were missed, and out of 720 significant stimuli presented under the experimental conditions a total of 14 signals were missed. It has already been pointed out that administration as an interpolated test probably showed the technique at a disadvantage. If our analysis so far is sound, we might reasonably predict that:

(a) An equivalent set of control scores would yield a number of missed signals not significantly different from zero.

(b) If the test were given during the heat exposure, we should obtain either a more striking effect for a similar set of conditions, or a similar effect with a less stressful set of conditions.

(c) Testing after a heat-plus-work condition should yield a significantly greater number of missed signals than testing after a heat-only condition of similar duration.

(d) If the test were given after, say, one hour to exposure to heat stress, and again after two hours of exposure, the number of missed signals in the latter case should be significantly higher than in the former.

These hypotheses, among others, were tested in the Farnborough Experiments, and the results follow.

FARNBOROUGH EXPERIMENTS

(a) Control Test

After practice, including dummy runs without heat (these were given for each fresh group of subjects), 24 subjects were tested three times before heat exposure, giving a total of 720 significant stimuli. As it happened, 2 signals were missed out of 720, but this is not significantly different from zero.

(b) <u>Exposure to increasing heat stress.</u>

It was not possible to obtain 100% saturation at the temperatures used for the Doncaster experiments. Experimental conditions, therefore, were as follows:-

(i) The mildest condition comprised temperatures ranging from 90 to 97.5 deg. (90% saturation) for 85 minutes under resting conditions. No other measurement was made, and no subject found the heat exposure unduly distressing. Nevertheless, an examination of the data shows that 9 missed signals occurred out of 720 presented under the experimental conditions. Applying once more the tests appropriate to a Poisson distribution, this proved to be significant at the 2% level.

(ii) Another group of eight subjects was subjected to a more stressful condition - exposure for a period of nine consecutive days to heat (dry bulb temperature 96.8-97.7 deg., wet bulb temperature 94.1-97.3 deg., air movement 165 ft./min.) for 120 minutes each day, under resting conditions. When tested after 120 minutes, a total of 14 missed signals occurred out of 580; this is significant at the 0.1% level.

(iii) The same group, after acclimatisation, was tested with a slightly more stressful condition (varying temperatures from 98.6/94.3 to 104/99.5, air movement 165 ft./min.) under resting conditions. The total number of missed signals was 9 out of 320 presented when tested after 120 minutes. This is significant at the 0.1% level.

(c) Heat-plus-work versus heat-only condition.

Finally, the same group was tested under the same conditions as in (b) (iii) but walking continuously on a treadmill at 3 m.p.h. Under these conditions, some of the subjects did not last until the end of the second hour. However, a comparison of the scores for equivalent duration of this and the previous conditions was possible, and this showed that 15 missed signals occurred during the heat-plus-work condition as compared with 6 missed signals for the comparative heat-only period. This difference is significant at the 2.5% level.

(d) First hour versus second hour.

The scores under the (b) (ii) and (b) (iii) conditions were pooled for the first hour of exposure and for the second hour of exposure, and compared. After the first hour 11 missed signals occurred out of 830 significant stimuli presented, and after the second hour 23 missed signals occurred out of 900 presented. This difference is significant at the 5% level.

It will be seen that the results are in good accord with the predictions.

HEAT STRESS AND REACTION TIME

Rather unexpectedly, the Farnborough experiments showed no significant change in reaction time to the significant signals throughout all the heat stress conditions including the heat-plus-work condition. In view of the earlier findings in vigilance experiments that vigilance reaction time is very much greater than the classical values, (Mackworth, 1950), this raises some interesting questions about the nature of the central processes involved in vigilance.

The findings from the present investigation suggest strongly that the probability of a missed signal during exposure to heat stress is small, but finite and constant for a given degree of stress. The reproducibility of the effect under markedly different conditions of administration appears to be high. But there is no significant change in vigilance reaction time, measured under these conditions. It may be that there are in fact two kinds of vigilance:

(a) Vigilance$_1$ - which occurs when a subject is set to observe a very occasional signal during a long period of watch. . In this case, we might expect the typical deterioration after half-an-hour and the greatly prolonged reaction times shown in the Clock-Test experiments and the earlier work on the Cambridge Visual Vigilance Test.

(b) Vigilance$_2$ - where the subject is set for a short period of observation, and the stimuli are presented at high speed and in comparatively large numbers. Here we might expect the small but finite and constant proportion of missed signals observed in the present investigation to occur with no change in the reaction time.

We know little as yet about possible underlying bases for these types of vigilance. But it may be that in Vigilance$_1$ there is a cumulative and comparatively gross reduction in receptor-effector efficiency, and that the greatly increased reaction time is an index of this reduced efficiency. In Vigilance$_2$ situation there may be a momentary failure on the part of the cerebral analysing mechanism, which is evidenced only by a small, constant probability of missing signals. It might not necessarily be shown in a mean reaction time score to observed signals, since the reduced efficiency is momentary and quickly made up again.

If this theory is tenable it should perhaps be possible to demonstrate an increased reaction time under heat stress when the subject is presented with a large number of signals in a comparatively shorttime, all of which must be reacted to. To test this a final experiment was set up where the subject was presented with a large number of stimuli in a short time (40 stimuli in 5 mins. in random order) and the mean serial reaction time before and during exposure to the most severe heat-less-work condition of the series was calculated. The data shown in Table 3 indicate that there is a small, but statistically significant, rise in this serial reaction time from the control period to the first hour; and from the first hour to the second hour.

TABLE 3

SERIAL REACTION TIME UNDER HEAT STRESS.

	Mean Serial Reaction Time	P
Control	0.247	
First Hour	0.256	
Second Hour	0.267	
Control v First Hour		0.05*
Control v Second Hour		0.01**
First v Second Hour		0.05*

* Significant

** High Significant.

It may seem surprising that such small differences should be statistically significant, but the standard error of the serial reaction time for a well-practised subject is remarkably small, about 0.0038.

DISCUSSION

Three practical conclusions emerge:-

(a) When an operator has to respond to a small number of critical signals occurring at irregular intervals and in an unpredictable order, his readiness to respond to the important stimuli will tend to suffer under heat stress.

(b) This deterioration in vigilance takes the form of a small but finite probability of missing a very clearly discriminable signal.

(c) As the degree of heat stress increases, the probability of a lapse of vigilance will also increase.

The undramatic nature of the effect may tend to obscure its significance. But this short-term vigilance effect is likely to be of critical importance in circumstances where the operator is required to respond immediately to occasional unpredictable signals coming in at high speeds. This is now the case in a number of Service situations, and in some industrial operations also. It would be interesting to examine a large number of well-documented accident records for situations where vigilance is required under varying ambient temperatures, and to look particularly for evidence of very occasional accidents occurring at temperatures normally considered to be reasonably comfortable, in view of the small but significant result obtained under condition (b) (i).

As far as recommendations with respect to the permissible time allowed at these various temperatures, it is not possible to give a set of direct values from the Doncaster data alone, since in fact, the figures quoted refer to the recovery period rather than to the period of heat exposure. It seems clear that some decrement in vigilance must be expected if members of the rescue team are exposed to any degree of heat stress, but presumably this has to be accepted. The important point to decide here is the level of deterioration in vigilance which is acceptable. In the absence of any other criteria it seems best to choose a cut-off point below which the number of missed signals expected for a given time will be less than 5%.

Examination of the Farnborough data shows two exposure periods which are comparable to the 95°F and the 100°F (light work) periods. The work in this case (walking on a treadmill at 3 m.p.h.) gives a mean energy cost of 155 Kcals/M^2/hr. which is fairly close to the value of 171 Kcals/M^2/hr. of the initial work period of the Doncaster trials. Considering first the equivalent temperature of the 100°F period, we find that there is a sharp rise in the number of missed signals from zero to about 12%, approximately 20 minutes after entry to the heat chamber. We can therefore recommend that not more than 20 minutes should be worked at the 100°F temperature of the rescue trials. Considering the temperature equivalent of 95°F in the Doncaster trials, we find that up to 25 minutes the percentage of missed signals is no higher than about 2%. Between 26 and 30 minutes after entry into the heat chamber there is a sudden rise to over 8%. We can therefore reasonably recommend that 26 minutes should be the maximum recommended time for the 95°F saturated temperature of the Doncaster trials.

There are no comparable data for the 90-925°F period but examination of the marked carry-over effect evident in the Doncaster data suggests that it would be unwise to recommend an exposure period of more than 35 minutes in view of the possibility that the heavy work involved after this period may bring about a very rapid deterioration in vigilance. Allocating this value to the 90°F temperature, to be on the safe side, we can interpolate a value of 30.5 minutes to correspond to the 92.5°F temperature.

We thus have four recommended time limits corresponding to the four highest temperatures. These are tabulated below (Table 4).

TABLE 4

Recommended Time Limits For Rescue Work In Temperatures Of 90°F, 9.25°F, 95°F And 100°F

Temperature	90°F	92.5°F	95°F	100°F
Time (Mins.)	35	30.5	26	20

It is scarcely practicable to give any figures for the Group 1 period with confidence. In view of the significant effect observed during the recovery period it seems certain that the period of exposure should be markedly less than the two mean tolerance times of 66.5 minutes and 64 minutes.

In view of the persistence into the recovery period of the vigilance decrement for at least 15 minutes, it would be extremely useful to know the permissible alternation of periods on and periods off duty. Mackworth's (1950) experiments showed that half-an-hour off duty is sufficient to restore full vigilance in radar operators, but it is highly questionable if this finding applies to rescue conditions where the physical strain is much greater than anything which a radar operator has to endure. Further research in this field is clearly desirable.

A full study of the interaction of temperature, duration of exposure and work done would require a long and carefully planned investigation. It is suggiciently clear, however, that there may be important differences in psychological reactions to variations in the time-temperature-work complex even where the end physiological conditions of the subjects appear similar. A complete mapping of the concomitant range of variation of psychological and physiological changes under the same stress conditions would be a most valuable contribution to scientific knowledge.

RESUME OF THE RESEARCH

1. The first section propounds the theme of the study - If vigilance is more than just a convenient label for a certain group of situations, it should be possible to demonstrate changes in vigilance occurring as the result of conditions extraneous to the vigilance tasks and not necessarily contemporaneous with it.

2. After a review of previous research on vigilance, the considerations which led to the development of the "z-function" (variability of judgment under vigilance conditions) test are described. The main feature of this vigilance task is that the subject has to make a series of estimates of the time at which a moving object reaches the centre of a projected display; the significant stimuli are interspersed randomly throughout a long series of neutral stimuli. The performance is scored by calculating the variance of the operator's estimates about his own mean estimate.

3. In the third section it is pointed out that the study is chiefly concerned with experimental work in the field to examine the changes in vigilance which occur under certain forms of stress. The nature of field experimentation is briefly discussed, and, since for much of the study the vigilance task is to be used as an interpolated test, certain problems involved in the use of interpolated tests are mentioned.

It is concludedthat there is no evidence to show any consistent effect of experience on the performance of vigilance tasks, although one investigator has suggested that individuals who show a marked improvement from the first to the second trial are more "fatigueable" than those who do not show this.

4. The fourth section deals with the validation of the technique in the field.

The results of the testing so far show a marked superiority in performance of the navigator subjects over the naval ratings, tested in previous investigations. The rise in the variance score after half-an-hour found in earlier investigations was confirmed in this experiment, as was the writer's earlier findings that reversal of this trend tends to occur in subjects showing a poor overall performance.

Scores on the test showed a significant deterioration effect after long sorties when compared with scores obtained from fresh subjects; no significant effect was found after short sorties. An individual case of reported "flying fatigue" was tested, and showed all three criteria on the test. Performance under stress conditions in the case of six of the best subjects showed a very marked deterioration after flying, but certain problems were encountered in the apparatus and scoring.

The results of a small number of cases suggested that the effects of flying a long sortie outweigh those of temporal position - i.e. time of day at which the test is given. Significant reversal effects were found after flying. The data obtained did not confirm the Welford - Brown - Gabb effect, viz that "measurable impairment of performance is likely to occur at tasks met for the first time when fatigued, but that little or no impairment occurs when the tasks have been met previously in a non-fatigued state". However, the tasks used by Welford et al were fairly complex skills, and the periods of flying covered were much longer, so that a cumulative fatigue effect might have been built up. More data would be required to judge how far the effect depends on the complexity of the task used and the duration of flying time.

5. The fifth section considers in some detail the definition and measurement of fatigue. The definition of fatigue has caused much controversy and there is still no general agreement; some writers consider that the term should not be used at all, while others would identify the term with the subjective experience of the individual. The present writer's opinion is that "fatigue" should be given an objective definition, reserved for situations in which some disorganisation of receptor-effector co-ordination can be demonstrated.

The measurement of fatigue has been attempted in two ways - by studies of deterioration during the task itself or during a similar synthetic task and by studies with tests given after "fatiguing" activities (interpolated tests). Studies of deterioration during the task itself have proved fruitful, but interpolated tests have proved generally unsatisfactory. The writer's submission is that this may be due to faults in the design and administration of such tests, one problem being that the novelty value of an interpolated test tends to mask a real deterioration. Applying certain principles derived from a study of "stress" experimentation, it is postulated that an adequate test for fatigue should -

(a) Involve continuous serial performance.

(b) Be applied under "stress" conditions.

(c) Provide minimal knowledge of results.

6. The next section discusses an investigation comprising nineteen 15-hour sorties, eleven by day and eight by night. One member of each crew was tested by the writer to obtain a measure of variability of judgment ("z-function"). The same individuals were tested when fresh to obtain a control score.

The results show:-

(a) that there is a significant difference ($P = 0.03$) between the control and experimental scores for the day-flying group, and a more significant difference ($P = 0.007$) between the control and experimental scores for the night-flying group.

(b) The increase in z-function for the night-flying group is significantly greater ($P = 0.01$) than for the day-flying group.

(c) The variances of the experimental scores in both cases are significantly higher than those of control scores.

These results are interpreted as indicating:-

(a) that a 15-hour flight in this type of piston-engined aircraft is liable to produce fatigue in the members of aircrew involved;

(b) that the fatigue involved in night-flying is greater than in day-flying of equivalent length;

(c) that increasing stress tends to produce increasing variability of judgment both within and between subjects;

(d) that the significant rise in z-function obtained after short night flights in the earlier experiment was not due to the fact that subjects were tested at a time of day when a fatigue effect would be expected in any case.

7. The next section discusses an investigation on long-range piston-engined aircraft to study the effects of cumulative fatigue as the result of flying four 15-hour sorties by night with one day's rest between each. It was concluded that:-

(a) Post-flight vigilance after the fourth sortie is significantly lower than after the first sortie. There appears to be some tendency for a progressive decrease in vigilance to occur as we go from the first to the fourth sortie, but this is not demonstrated as significant in the present experiment.

(b) Flying four 15-hour sorties at night with a day's rest between each sortie is likely to produce considerable fatigue in the crews concerned, manifested both as objective deterioration in post-flight vigilance and as subjective reported weariness and exhaustion.

(c) Subjective fatigue and objective deterioration in vigilance are significantly correlated in this investigation.

8. Since in this investigation a number of other variables, both performance and physiological, were studied, an attempt is made to predict which of the measures are most likely to show effects of fatigue. The results indicate some support for the anglysis made in section 5, although differing times and motivations make it difficult to compare the various measures. One thing which comes clearly out of the study is that significant performances changes can be demonstrated without any definite change in most of the physiological measures employed.

9. In the next section we return to an earlier finding. In the jet-engined fighters we have a fatigue effect developing after 2-3 one-hour sorties in a small smooth-flying, quiet aircraft involving

little effort to handle and unaccompanied in many cases by subjective reports of weariness. In the piston-engined aircraft studied we have a fatigue effect developing after a continuous flight of ten hours or more in a large, heavy, noisy aircraft involving a fair degree of physical effort to handle and accompanied by marked subjective weariness in the crew. It seems advisable to analyse the jet situation further to see if we can gain more information about the fatigue inducing variables.

10. The next section, therefore, describes two further investigations. In the first, aircrew were tested before and after flying for a 3-4 hour continuous sortie in a jet bomber carrying out a routine excercise. No significant fatigue effect was found, and changes in either direction were small. In the second investigation, a small group of pilots flying jet-engined fighters were tested under carefully controlled conditions, before and after flying 3 one-hour sorties. All subjects showed a rise in z-function after flying, and this rise was significant at the 2% level.

These experiments make it clear that jet flying per se is not necessarily more fatiguing than flying in piston-engined aircraft. A final analysis of how far the fatigue effect depends on the multiplication of sorties awaits the development of jet fighters with long-range fuel tanks.

11. So far the experiments have been entirely on flying fatigue, using the z-function measure. The next section discusses the possibility of reducing the length of the test to facilitate the testing of more subjects. One way to do this is to speed up the display so that the subject has the same number of significant signals in, say, fifteen minutes. This was tried in two classes of fatigue situation,but no significant results were obtained. Presumably the testno longer involves a stress condition, and therefore, according to the analysis of section 5, is not likely to function as an interpolated test of fatigue. A new form of task (the Cambridge visual vigilance test) was devised, and used in the subsequent investigations on heat stress. The primary objeets of this were to check that a straight vigilance task could be used as an interpolated test for effects of stress in the same way as the z-function test, to try the effects of a test which involved shorter duration but much higher speed of presentation, and to examine how far vigilance is affected by another form of stress condition.

12. In the next section, an experiment is described in which this Cambridge visual vigilance test is given to members of mines rescue teams before and after exposure to heat under conditions resembling those of rescue work during mine accidents.

11. The results showedthat exposure to heat stress of this order produces a significant deterioration in performance of the task when all the results are pooled.

The number of missed signals occurring during the experimental conditions was significantly increased above the control level (zero) after exposure to temperatures of 85-87.5 degrees and increased still further after exposure to temperatures of 90-92.5 degrees; for temperatures above this level the number of missed signals did not differ significantly from the control score.

The simplest explanation of this appears to be that deterioration of vigilance is a function of (a)temperature, (b)duration of exposure, and (c) work done; at the lower temperatures the subjects remained in the chamber for the full time, while at the higher temperatures they were withdrawn earlier, thus missing the heaviest part of the physical work.

An additional test involving subjects exposed to a temperature of 100 degrees and given heavy work from the beginning of the exposure produced no increase in missed signals but showed a marked decrease in the number of fast reactions; this, however, was not significant for the number of subjects tested.

13 A further series of experiments was carried out to test the hypotheses that:-

(a) An equivalent set of control scores would yield a number of missed signals not significantly different from zero.

(b) If the test were given during the heat exposure, there should be obtained either a more striking effect for a similar set of conditions, or a similar effect with a less stressful set of conditions.

(c) Testing after a heat-plus-work condition should yield a significantly greater number of missed signals than testing after a heat-only condition of similar duration.

(d) If the test were given after, say one hour of exposure to heat stress, and again after two hours of exposure, the number of missed signals in the latter case should be significantly higher than in the former.

The results were found to be in good accord with these predictions.

No significant change in vigilance reaction time to the significant signals was observed throughout all the latter heat stress conditions, including the heat-plus-work condition.

It is therefore suggested that there may be two kinds of vigilance:-

(a) Vigilance$_1$ - where the subject is set to observe a very occasional signal during a long period of watch, and (b) Vigilance$_2$ -

where the subject is set for a short period of observation and the stimuli are presented at high speed and in comparatively large numbers. If this is true, it should be possible to demonstrate an increased reaction time under heat stress where a subject is presented with a large number of stimuli in a short time.

To test this latter hypothesis, subjects were presented with 40 significant stimuli in five minutes before and during exposure to the most severe heat-less-work condition. The results showed a small but statistically significant rise in serial reaction time during heat exposure.

SUMMING UP

One main object of the study was to show that suitable tests can demonstrate changes in vigilance occurring as the result of conditions extraneous to the task itself, and not necessarily contemporaneous with it. The experiments have shown significant deterioration in vigilance as the result of, or, at least, in association with:-

(a) long-range flying in piston-engined aircraft

(b) repeated short sorties in jet-engined fighters

(c) exposure to a certain range of ambient temperatures without physical work

(d) exposure to a certain range of ambient temperatures with physical work.

Moreover, the evidence is consistent with the hypothesis that deterioration of vigilance begins to set in after a certain level of stress is applied - i.e. after a certain duration, above a certain temperature, after so many sorties, etc.

A second object of the study was to show how well a vigilance task will function as an interpolated test for fatigue. The experimental work indicates that it is more effective than most tests of this kind. In the theoretical analysis of the reasons for this, the writer has postulated that the vigilance task fulfils three main conditions for a satisfactory fatigue test, since it involves continuous serial performance, is applied under "stress" conditions, and involves minimal knowledge of results.

While a full test of these criteria would require very elaborate investigation, we have a certain amount of supporting evidence from the fact that (a) a reduction in the time of application of the z-function test rendered it ineffective as an interpolated test, and (b) increase in the speed at which discrimination had to be made apparently renders the vigilance task more sensitive as a test for fatigue.

Much of the earlier work on skilled performance has demonstrated the importance of knowledge of results in enabling the operator to maintain a certain level of performance under stress or fatigue (Bartlett, 1948). It is clear that knowledge of results interpreted in a very broad sense plays an important part in the maintenance of performance in vigilance tasks. Mackworth's (1950) data suggest that supplying knowledge of results is enough in itself to prevent the typical pattern of deterioration in the second half-hour of the clock test experiments. It seems that what this kind of task does is to present the individual with a relatively structured situation in which certain aspects of the performance can be precisely measured, but in which he is given no clear index as to how his performance is being maintained. By thus imposing less constraint on the subject and allowing a freer operation of motivational factors, we are coming perhaps closer to the real-life situation than the more standardised laboratory tasks do, to studying the gap, as Bartlett (1948) puts it, between what the operator can do and what he will do.

The effectiveness of the vigilance task as a measure of stress is corroborated by the fact that Broadbent (1951) has been able to show a significant effect of ambient noise (100 db.) on performance of one of his vigilance tasks; this, according to Bartlett, is the first experimental proof that noise has a fatigue effect.

A third object of the study was to test the hypothesis that increasing stress tends to produce increasing deterioration in vigilance. This is to some extent confirmed by the heat stress studies, and, to a lesser extent, by the flying fatigue experiments. Again, a full test of the hypothesis would require a very elaborate investigation, but the data here support and supplement earlier work of the writer's (Fraser, 1951) and other investigators (Mackworth, 1950; Conrad, 1951).

The relation of the data provided here to other work in the field of experimental psychology and physiology has also been discussed during the presentation of the research. Certain additional points may be worthy of consideration here.

The emphasis on variability of performance as an indication of fatigue or stress was developed from the writer's stress theory, outlined in section 5, and is reflected in the method of scoring of the z-function test. Two recent studies have also examined this aspect of performance. Conrad (1955) has shown in a factory task that the mean output does not change as the day goes on, but that the variability of performance of the workers rises significantly. Siddall and Anderson (1955) have shown, in a study of fatigue in a continuous tracking task, that both the means and the standard deviations of the number of errors, and the duration of errors rise steadily with "fatigue".

It may be useful to consider briefly the significance of measuring variability in the vigilance task. If we devote our attention to the missed signals in a fixed task like Mackworth's clock test, we appear to find that vigilance falls off markedly after the first half-hour and then remains at the same level approximately for the next one-and-a-half hours. Now, this is a feasible result, but it does not quite fit the picture of deterioration obtained with other forms of stress. Also, Mackworth records an experiment with the synthetic radar apparatus, in which echoes of two different intensities were used. When the brightness of the echo was at low intensity, the typical deterioration appeared in the second half-hour. With the increased intensity, the deterioration did not become statistically definite until one hour after the start of the spell. The present writer, moreover, has been able to show that alterations in the ratio of the stimuli used in the earlier form of the vigilance task described in Section 2 tend to increase or decrease the "fatigue" effect appropriately (Fraser, 1951). As a result of these considerations, the writer had advanced the theory that deterioration of vigilance tends to occur at certain critical levels of, e.g. stimulus ratio, intensity, duration; but when these critical levels are exceeded, vigilance may still continue to decline as the stress increases (Fraser, ibid.). It is, in fact, suggested that if we choose a certain combination of levels of intensity, etc. and use a gross criterion like number of signals missed we can level out the variability of performance which would appear with a more sensitive and continuous criterion.

It is now appropriate to consider how far the experiments discussed have contributed to our understanding of the concept of vigilance, in addition to answering the more specific problems raised at the beginning of the research.

In spite of the not inconsiderable amount of research which has been done in the field of vigilance there are some features about the work which cause confusion in discussion. This is well brought out in three recent articles by Bakan (1955), Deese (1955) and Elliott(1956). Bakan and Deese devote much of their attention to defects in Mackworth's (1950) techniques and theoretical analysis, while Elliott examines the problem on a wider basis. It is perhaps worth listing at this point the criticisms which can be made of the clock test as a vigilance test, since it was the pioneer of the studies in this field.

In the first place, the clock test involves the possibility of an auditory cue to the appearance of the significant signals.

Again, the clock test involves the repetition of the same pattern of signals every half hour. This is subject to criticism from two angles. The pattern of occurrence of significant signals may be learnt, as occurred with certain of Pepler's subjects. This learning was evidenced by a striking improvement in performance after a certain number of trials. It may also be that some degree of learning occurs in other cases, although the improvement in performance is not so dramatic. Clearly this is an undesirable variable in the experimental situation.

Another criticism of this arrangement (made by Bakan) is that in the practice period a specific temporal expectancy is built up during the first five minutes, and that this expectancy is confirmed during the first half hour but does not operate so effectively over the second, third and fourth half hours. It would be possible, of course, to meet this criticism by dispensing with or diminishing the length of the practice period.

A related point which has already been made, although it is not necessarily a serious criticism of the clock test results, is that the relation between the neutral and significant stimuli cannot be altered to any considerable extent; this makes it a rather inflexible medium for laboratory investigation.

It will be apparent however that none of these criticisms is applicable to the techniques used in the present researches.

When we come to examine the theoretical explanations of vigilance, we also find some disagreement. Bakan, for example, questions the necessity for bringing in the framework of a laboratory conditioning situation, since the response could be set up by telling the subject what to do. Again, he considers that the rejected hypothesis of conditioned inhibition is too specific for most vigilance situations. It is quite likely that knowledge of results might forestall the secondary extinction or conditioned inhibition also. He suggests the desirability of a discrimination orientation, the underlying hypothesis being that the ability to discriminate declines during work at a task requiring prolonged attention - in other words there is a progressive increase in any appropriate measure of the differential threshold.

The nature of this change in differential threshold might take any one of three possible forms:-

(a) there is a decrement which is independent of the intensity of a stimulus. In other words, there is a lowered likelihood of response at all stimulus levels.

(b) the psychometric curve shifts up the stimulus scale. In other words, there is a broader zone of equality, with differences beyond this zone reasonably sharp.

(c) there is an increase in the stimulus range over which the psychometric function extends. In other words, there is a less sharply defined "equals" category after decrement.

This analysis is compared with the writer's later in the discussion.

Deese would prefer to substitute for Mackworth's concept of an inhibitory process the basic assumption that the maintenance of a given level of vigilance in the observer depends partly on stimulus events extrensic to the observer. The main emphasis is therefore on the excitatory properties of vigilance, ignoring the assumed inhibitory properties. He considers the possibility of a reinforcement hypothesis in which it is assumed at the outset that the observer operates at some specified level of efficiency and that this level is constant from one period of search to another. It is assumed that this probability of detection goes through a decline after an initial warm-up period, this decline representing a dissipation in the initial excitatory state of vigilance. The critical assumption here is that the occasional significant signals which occur in the observer's field of research will determine the future course of his ability to detect: Specifically, either the occurrence or detection of a signal would produce an immediate return to the initial level of vigilance or to a level determined by change in other variables.

It certainly appears to be the case that detectability is higher when signals appear more frequently. Deese and Ormond (1953) showed an increase from 47% to 87% of detected signals when the number of signals per hour was increased from 10 to 40. This finding is confirmed by some early work of Broadbent (1951) and some unpublished experiments of Jenkins (1954) and Bowen (1956).

We might, on the other hand, consider that detectability is a function of the size of the intersignal interval rather than the rate of signal appearance. The re-inforcement hypothesis would predict that detectability of a second signal would be greater the smaller the time interval between two signals. But the data from Deese and Ormond (1953) do not support the hypothesis. The intersignal interval either has no effect on the probability of detection (within the given rate) or the effect is in the opposite direction to the prediction. So Deese formulates another hypothesis - an expectancy hypothesis. The main points of this are:-

(a) the observer's expectancy or prediction about the search task is determined by the actual course of stimulus events during his previous experience with the task.

(b) the observer's level of expectancy determines the vigilance level, hence the probability of detection.

This gives the same prediction with respect to detectability and average rate of signal appearance as that made by the reinforcement hypothesis. But the probability of detection for a given signal in the series is the same as for any other signal, assuming that the distribution is not some non-random function of time.

This assumption in fact implies a linear relationship between the average of some selected aspect of past events and vigilance with respect to events at any given moment. If the task is invariant with time, vigilance will change little after an average level of expectancy has been reached. Such an assumption demands that the observer's activities should be simpler than they probably are.

It is reasonable to assume some short term fluctuations in expectancy. On this basis, if for any given signal the observer's expectancy is exactly determined by the mean intersignal interval, and if his ability to estimate time since the last target is free of constant error and shows relatively low variable error, the expectancy should be low immediately after a signal, should increase as the mean intersignal is approached, and finally should become quite high as the intersignal interval grows beyond the mean. The mean probability of detection averaged over these periodic fluctuations would correspond to the average intersignal interval.

The actual evidence available from Deese and Ormond and Mackworth's experiments provide some indication that the probability of detection increases with increasing intervals between the presentation of signals. But the overall picture is not very clear in this respect.

Elliott (1956) also invokes the term expectation. If the operator expects a signal every second or two and in fact gets signals at that rate, then his performance is likely to be optimum. On the other hand, if the operator expects to get signals rarely and in fact gets them rarely, it might be anticipated that his performance would be poorer than optimum from the moment he forms this expectation. In an operational search task and in a laboratory task where the operator has had previous experience of it and where the instructions are adequate to form a correct expectation that signals will appear rarely, the operator's expectations are formed before the test starts, and therefore the performance from the beginning of the task is likely to be worse than optimum.

It is rather difficult to see how this concept of expectancy can be treated in any rigorous way. Clearly if the operator expects signals to occur at a certain point in time then we may reasonably expect him to watch out particularly for them and therefore presumably to be rather more successful in the test. But apart from a very broad assumption of this kind the "hypothesis" appears to have rather little predictive value. Moreover, it is difficult to explain why knowledge of results in the original experiments of Mackworth appears to have such marked infuence on subsequent performance.

Returning again to the question of methodology, both Bakan and Elliott have devised a technique for measuring vigilance, in which an undetected signal presented near the threshold level is presented again at successively higher strengths until it is detected. This method, or a similar one. has also been used by Bowen in some unpublished work on vigilance. These observers consider that this technique is more satisfactory than Mackworth's in giving an estimate of the rise in threshold which occurs as the observer's vigilance decreases. Bakan (1952) found an increase in threshold during the first half hour of the test followed by a tendency to level off for the rest of the period. The decrement was such that a brightness increase of about 25% late in the vigil was required for detection. This appears to agree quite well with Mackworth's original findings. Elliott, on the other hand, finds considerable fluctuation in vigilance.

It may be argued that these techniques are not as satisfactory as their inventors believe. If there is any truth in the hypothesis that the probability of detection increases as a function of the number of successful detections, then this method of ensuring that the subject eventually has knowledge of results may make the task different in certain important respects from the true watch-keeping situation. Undoubtedly the technique is measuring some aspect of vigilance, but it is questionable whether it is possible to generalise from the findings on this as legitimately as from the more orthodox task.

Inn addition, it may be that the repeated presentation of signals of increasing strength results in some form of sub-threshold perception. There does seem to be some ground for believing that unrecognised detection of signals may occur, in fact, Deese's hypothesis is framed to allow for either the occurrence or the detection of a signal affecting the probability of detection of subsequent signals.

From the consideration of the arguments above one can draw two general conclusions:-

In the first place, it is necessary to distinguish between the terms "vigilance" and "vigilance tasks". "Vigilance" refers to a state of alertness or readiness to respond to significant changes in the environment. It has the status of an intervening variable, and we can really only speculate about the central factors which mediate it in the present state of research. A "Vigilance task", on the other hand, is a clearly defined experimental arangement characterised by the three following properties;-

(a) The subject is presented with a display consisting of a series of neutral signals throughout which are interspersed at random intervals a number of significant signals.

(b) The conditions of the experiments are such as to render it a stress situation - i.e. the signals are presented at such a high speed or with such a heavy load or for such a long duration that there is a significant probability of deterioration during the task.

(c) The knowledge of results of the subject's performance is minimal. If, and only if, these three conditions are fulfilled, should we expect the findings of the early vigilance experiments to apply. In situations where these three conditions are observed there is a considerable measure of agreement as to the general characteristics of vigilance tasks. But if we alter one or more of these properties we cannot expect similar results from the experiments. So a set-up, like that of Elliott or Bakan, where an undetected signal is presented again at a higher strength, is not a true vigilance task, by definition.

Secondly, it is commonly assumed that breakdown in vigilance is of the same kind whatever the watch-keeping situation. But, as the results of experiment 6 of the present study have shown, there are probably at least two patterns of breakdown, and possibly therefore two kinds of vigilance:-

(a) Vigilance $_1$ - Where the subject has to watch for an occasional significant signal appearing amongst a large number of neutral signals over a long period and with comparatively slow presentation. In such a case, as we have seen, we may expect the typical falling off of the early Mackworth, etc., experiments.

(b) Vigilance $_2$ - Where the subject is presented with many signals appearing only for a very short time and the pattern of breakdown is intermittent and very quickly compensated for. Such a form of breakdown would not necessarily show the same features as the typical vigilance situation.

In discussing the implications of this theory earlier, the writer suggested that one effect of stress was to cause occasional momentary failure of the cerebral analysing mechanism. This momentary failure has probably been overlooked in the past because it is rapidly compensated for, and would only lead to the omission of response if the signal calling for action appeared simultaneously with the failure of the cerebral mechanism, a signal which was not prolonged until the recovery of the analysing mechanism. If in normal operations the distribution of the incoming signals tends to be regular, then the operator can anticipate the occurrence of the next signal; the human central nervous system appears to be well adapted to responding to such regular sequences. If, on the other hand, the distribution of signals is irregular in time, we should expect at some point coincidence of the cerebral analysing mechanism failure and the occurrence of a significant signal. If the occurrence of the signals is comparatively infrequent, and the duration of the individual signal is very short, there should be a small but finite and constant probability of missing a significant signal. The smallness of this probability, measurable perhaps only in the form of a Poisson distribution, almost certainly accounts for its failure to be observed in previous investigation. Hence the importance of the tie-up between the Doncaster and Farnborough experiments of the present study since they show that lapses of vigilance under rather different sets of heat stress conditions and with two markedly different groups of subjects appear to be accurately predictable.

Another deduction which might be made from this theory is that some test of the efficiency of the cerebral analysing mechanism should show a deterioration as a result of stress. If we consider that reaction time is one measure of the efficiency of nervous transmission we should expect this to be impaired under stress. A point of interest here is that form of reaction time which should be used as a measure. A single measure of simple reaction time would obviously be quite unsuitable. A large number of discrete measures of simple reaction time with the warning signal presented at some optimum time before the stimulus would not be expected to show any change, in view of the intermittent nature of the presentation. If we presented the simple reaction time stimuli as a series, distributed irregularly in time and with the appropriate warning at some optimum time before the arrival of the stimulus, we might expect some measurable deterioration to occur. But, again, this arrangement would allow the operator to anticipate the occurrence of the significant signal, and, according to our theory, the central nervous system should be able to handle such a situation, even under stress.

The only effective technique, then, would appear to be a measure of serial reaction time, where a large number of signals, distributed irregularly in time, is presented to the subject without warning. As we have seen, serial reaction time does show a small, but significant rise after heat stress.

A point of considerable theoretical interest which now arises is the relation between Vigilance $_1$ and Vigilance $_2$. It has been suggested in this study that these are of two different kinds. It might be, however, that Vigilance $_1$ is an exaggerated form of Vigilance $_2$.

In the present stage of research, we have comparatively little evidence to settle this point, but it is interesting to consider the reasons why the two forms of vigilance should be disparate. Perhaps the main difference between the two situations is in the frequency of external stimulation. Recent research (Bexton, Heron, and Scott, 1954) suggests that the maintenance of normal adaptive behaviour probably requires a continually varied sensory input. In the Vigilance $_2$ situation the observer has a sensory input with frequent variations even if these are of a rather simple kind. In the Vigilance $_1$ situation on the other hand, the sensory input from the task itself is infrequently varied. Therefore, we might consider this point of view as supporting the theory of two different patterns of vigilance.

Now, if the patterns of vigilance differ according to the form of presentation, we should expect that the pattern of deterioration of vigilance occurring as a result of a long watch-keeping situation would in some respects be unique; the name "Vigilance Stress" has been suggested by the writer for this distinctive pattern of breakdown under stress.

If we discount the problems involved in the use of methods such as Elliott's and Bakan's, it appears that measurable deterioration in vigilance is probably a function of:

(i) The duration of the significant stimuli

(ii) The duration of the task

(iii) The frequency of signal

(iv) The difficulty of discrimination involved in the task

(v) The informational feedback inherent in the task

(vi) The degree of stress under which the task is performed

(vii) Individual differences between subjects

(viii) The environment of the subject

(ix) The present state of the subject.

If our theory is sound, it should be possible to make certain predictions from it about some or most of these factors:

(1) Duration of significant signals

The theory is quite specific here. Since variability of judgment and vigilance is held to increase as time goes on, a longer signal, late in the vigil, is more likely to strike a peak at some point in the fluctuating curve of vigilance.

Conclusion The longer the duration of the significant signals, the more likely they are to be observed.

This is borne out by some earlier experiments of the writer (Fraser, 1951), by the fact that diminishing the duration of the significant signals appears to make the vigilance test more sensitive to the effects of fatigue in the present study, and by some observations of Broadbent (1951, 1953) on the difference between transient and remaining signals.

(2) Duration of the task

The theory should predict increasing deterioration of vigilance with increasing duration of the task. This apparently conflicts with Mackworth's and Bakan's finding that deterioration is most marked after the first half-hour, after which it appears to settle down to a steady level. But, as we have already seen, this may be accounted for by the fixed level task used by Mackworth and Bakan. Moreover, it is not corroborated by the findings of Deese and Ormond, nor by Elliott. Unfortunately, the present study has not been concerned with really long duration, so that we cannot be very certain about this aspect of vigilance. If subsequent experimentation confirms the Mackworth type of result, it will be necessary to develop a satisfactory theory to account for this major decline after the first half-hour.

Conclusion The longer the duration of the task, the greater the likelihood of a decrease in vigilance.

The evidence in this respect is conflicting.

(3) Frequency of signal presentation

The theory does not deal very specifically with this factor, except that infrequent signals may be held to involve a lower degree of informational feedback than frequent signals. But no very precise conclusion can be drawn otherwise.

(4) Difficulty of discrimination

Within a certain range of difficulty we should expect that the predicted increase in variability of judgment of the subject would eventually mean that signals at a strength likely to be detected originally would be missed later in the task. This is a similar deduction to that made by Bakan which was discussed earlier, except that Bakan's discrimination orientation requires that a choice be made from three patterns of change in the differential threshold. The first of these possibilities - that decrement is independent of intensity - is contrary to common sense. It is also contradicted by the fact that an early experiment of the writer (Fraser, 1950) showed that there was a smaller decrement in vigilance when the double movements of the hand in the clock test were made easier to see, and by Mackworth's own observation that detection was better and decline of performance less steep when a brighter stimulus was used in the simulated radar test. Neither of the other two possibilities advanced by Bakan - that the psychometric curve shifts up the stimulus scale, or that there is an increase in stimulus range over which the psychometric curve extends - is quite comparable to the present hypothesis of increasing variability as time goes on, but both types of analysis would predict an increasing likelihood of deterioration in vigilance as the difficulty of discrimination is increased.

<u>Conclusion</u> Within a certain range, increase in difficulty of discrimination tends to increase the likelihood of a lapse of vigilance.

Confirmed by the evidence mentioned above.

(5) Informational feedback

Again, this is an integral part of the theory. The more complete, the more precise, the more immediate the knowledge of performance, the less the performance is adversely affected by stress, and this holds for vigilance stress.

Conclusion The greater the degree of informational feedback, the less the likelihood of a lapse of vigilance.

Confirmed by Mackworth's (1950) experiments, by some early work of the writer (Fraser, 1951), by Deese and Ormond (1953), and by some unpublished work of Wilkinson.

(6) Degree of stress

Another integral part of the present work. This appears to be an important finding, since it offers the possibility of a precise quantitative prediction of the extent of vigilance decrement under stress.

Conclusion The greater the degree of stress above a certain level, the greater the likelihood of a lapse of vigilance.

Confirmed by the experiments of the present investigation.

(7) Individual differences

Examination of the individual scores in vigilance tasks shows considerable differences between subjects in all forms of the task. It seems very likely that there are basic individual differences in vigilance, and it would obviously be satisfactory if the mean scores obtained by several runs of testing on a vigilance task reflected these individual differences with sufficient accuracy to allow us to make some definite predictions about the behaviour of these individuals in situations outside the laboratory. As we have seen, Broadbent suggests that individuals who tend to show a marked improvement from the first to the second trial are more "fatigueable" than those who do not show this trend, and has found some corroborative evidence for this in a significant correlation between the tendency to improve in the test situation and the frequency of change of occupation and lack of persistence in continuing with higher education. If this is confirmed by further investigations, it would throw a good deal of light on the significance of these individual differences.

The writer attempted in the initial stages of the aircrew fatigue investigation to obtain a rating from the C.O.'s of the squadrons from which subjects were selected which would give some indication of personality characteristics of this kind. A seven-point scale for rating was put forward, so that some kind of comparison between different raters could be made. The main difficulty, however, was that of obtaining a satisfactory spread of ratings in the larger squadrons. In the case of the small, compact fighter squadrons, where the crew of each fighter consisted of one or two officers, most C.O.'s were prepared to give ratings over almost the whole of the scale, but in the larger reconnaissance aircraft squadrons, where there might be a dozen aircrew in each aircraft, the C.O.'s tended to play safe, and give the majority of the the ratings around the average score.

In the case of the fighter squadron ratings, it appeared that no subject who was given a really good rating for keenness, application and industry ever scored really badly on the control tests with the vigilance task in use. But, unfortunately, several subjects who scored well on the vigilance task were given poor ratings for the above qualities by their C.O.'s, so that any hope of a simple correlation between the two was dispelled. However, it seems much more likely that a series of tests, given in a controlled and exacting experimental setting like that of the cumulative fatigue investigation would offer more hope of a significant correlation with personality traits of the kind related to fatigueability.

Of course, the present theory does predict a connection between individual differences in skill and the stage at which breakdown of performance begins; but it would be stretching this considerably to extend it to individual differences in vigilance unless it proved possible to demonstrate that measured vigilance is actually an index of psychomotor efficiency, in much the same way that Head considered neural vigilance to be an index of physiological efficiency.

In general, it appears that no very specific prediction can be made from the theory so far about individual differences in vigilance. However, research into this aspe ct of vigilance is clearly of great importance, both from the practical angle of detection, and for the light which it might shed on the underlying cerebral concomitants of vigilance.

(8) Environment of the subject

If, as is suggested in the earlier part of this discussion, the conditions of administration of a vigilance task render the test itself stressful ("vigilance stress"), supplying more sources of variation in the environment, provided these are not so marked as to distract the subject, should help to maintain the level of vigilance.

Conclusion A homogeneous environment is more conducive to vigilance decrement than a varied one.

Confirmed by an experiment of the writer (Fraser, 1952), in which the presence or absence of the experimenter produced a reliable difference in the number of signals detected; this difference was contributed entirely by seven of the eighteen subjects. A further examination of this type of result should throw additional light on the nature of personality differences in the field of vigilance.

In a general way, such a conclusion might also be inferred from the Bexton, Heron and Scott type of analysis.

This conclusion offers a useful clue to the maintenance of vigilance, and has, in fact, been the basis of a number of suggestions by theorists interested in the optimum detection of subthreshold signals.

(9) Present state of the subject

This, in the writer's opinion, is an aspect of vigilance which has not been considered in anything like sufficient detail by theorists, and may in itself be enough to account for some of the disparities in research findings.

The theoretical analysis of the definition and measurement of fatigue on pp. 28-34 led up to the conclusion that fatigue was a disorganisation of receptor-effector co-ordination, resulting from exposure to high speeds, loads, duration, anxiety, conflict, sleep deprivation or other form of stress condition. Now, this disorganisation may be quite evanescent, as seems likely in the case of the speed stress type of experiment, where the breakdown in performance seems to be limited to the actual peak periods of speed stress (although even here there may be some carry-over in certain subjects; Venables (1954) has made this the basis of a personality test). On the other hand, it may be a more enduring disorganisation.

Perhaps the chief contribution of the present study has been the demonstration that exposure to a fairly high degree of stress, subjectively established, can produce measurable changes in the performance of a vigilance task, not only during the stress exposure, but afterwards. The changes are not always very dramatic, but they are statistically reliable and appear to be quite reproducible. It is a matter of common experience, as Bartley and Chute (1947) emphasise in their discursive exposition on the importance of subjective aspects of fatigue, that a subjectively stressful experience seems to carry over into later, unstressful activity. But, perhaps for the reasons suggested in the earlier analysis of this problem, it has proved very difficult to study and measure this carry-over, and therefore very difficult to achieve agreement among investigators as to its severity and manifestation. If we agree that the present evidence, backed by additional research which is now going on in Cambridge, Farnborough and Oxford, establishes the validity of this theory that high speeds, loads, anxiety, conflict, sleep deprivation, etc. may produce changes in the functioning of the central nervous system which continue, as it were, to reverberate in some form throughout the c.n.s. in later activity and may affect[1] that activity in a predictable way, a number of important conclusions follow. (Emphasis must be laid on the word 'predictable', otherwise we are doing little more than confirming common experience by time-consuming investigation, an activity for which psychologists are often, if sometimes undeservedly criticised).

One important conclusion, and the one which is particularly relevant in the present context, is that we cannot assume that any given subject sitting down to begin a vigilance test is, in fact, a 'fresh' subject. This may seem an obvious statement. It is highly unlikely that all individuals, or the same individual on successive occasions, will approach a given psychomotor test with the same degree of freshness. But, if our analysis on pp. 28-34 is sound, the usual psychomotor test, often involving a high degree of informational feedback, is not likely to show the effects of antecedent stress, while the vigilance task is.

---

1 'Fatigue' is broadly distinguishable from learning in this context in that it tends to be associated with disorganisation rather than organisation, and to disappear after a time, but it is scarcely possible to draw and entirely satisfactory distinction.

Now we have already seen that the evidence in the Mackworth type of vigilance task points to a marked decrement in vigilance after half-an-hour, while performance in the second, third and fourth half-hours remains about the same level. Let us suppose that the test is administered to a subject who has been deprived of a night's sleep, and that the carry-over from this is sufficient to depress his level of vigilance to such a degree that he starts the vigilance task about the level which would normally be expected after half-an-hour on the task. The picture of performance which would result would be one of rather irregular operation from the start with no significant changes from one half-hour to the next. It is not suggested that this in itself is enough to account for the different findings of experimenters using different forms of measures of alertness, but it might be an important factor, particularly since some techniques appear to be more sensitive to fatigue of this kind than others.

There is some evidence to show that sleep deprivation does impair subsequent performance of a vigilance task. The writer has frequently found that testing subjects who have arrived at Farnborough after a night train journey, with rather intermittent sleep, produces a worse set of scores than usual, even for an initial run. Wilkinson, in some unpublished experiments in Cambridge, has been able to show a statistically significant change in vigilance after the loss of a night's sleep.

Moreover, the writer has found some evidence (Fraser, 1956a), although it is very tentative at present in view of the very stringent statistical control necessary to establish conclusions of this kind, that vigilance is higher in certain individuals at different times of day.

Drawing a general conclusion from the discussion of this and the previous subsection, it is clear that further research on vigilance could profitably be concentrated on individual differences between subjects, and the present state of the subject at the time of testing, and that such research might go a long way to ironing out the apparent inconsistencies in the findings of different experimenters.

It is not easy to round off a study of this kind neatly. The investigation has given a fairly definite answer to several of the problems which were raised at the beginning, and thrown some light on others not anticipated in the original framework of research. But one of the main difficulties in a rigorous analysis of both vigilance and fatigue is the lack of continuing research into their manifestations in a variety of field situations; the attempt to make up for some of this deficiency was a major aim of the present study, in its inception.

The present study also suggests that the research tools at the moment employed in this field are comparatively clumsy compared with those which could be devised to combine the best (i.e. most sensitive to fatigue) features of a fatigue measure. An obvious improvement would seem to be the addition of load stress to speed and vigilance stress elements in the measure, and a single apparatus designed to study five different parameters of performance at the same time is now being tested at Farnborough. It may be that the results obtained from this kind of measure will make the present investigation seem like a sledgehammer of effort to crack some nuts of information. But if further research confirms the soundness of the kernels, we should be some way to solving this critical problem of the nature and measurement of fatigue.

REFERENCES

BAKAN, P. (1953) Unpublished Report.

BARTLETT, F.C. (1948) The measurement of human skill occ. Psychol. 1948, 22, No. 1, pp. 31-38.

BARTLEY, S.H., & CHUTE, E. (1947). Fatigue and impairment in man. New York: Mc-Graw-Hill 1947.

BROADBENT, D.E. (1951) Noise, Paced Performance and Vigilance Tasks, 1951. A.P.U. Report 165. Applied Psychology Research Unit, Cambridge.

BROADBENT, D.E. (1953) Classical Conditioning and Human Watch-Keeping. Psychol. Rev., 60, 331-339.

CARMICHAEL, L. & DEARBORN, W.F. (1948). Reading and Visual Fatigue, Harrap & Co., London.

CONRAD, R. (1951) Speed and Load Stress in Sensori-Motor Skill. Brit. J. Indust. Med., 8, 1-7.

CONRAD, R. (1955) A Comparison of Paced and Unpaced Performance at a Packing Task. Occup. Psychol., 29, 1, 15-28.

DEESE, J., & ORMOND, ELIZABETH, (1953). Studies of detectability during continuous visual search. WADC Tech. Report, WADC-TR-53-8, 1953.

DEESE, J. (1955) Some Problems in the Theory of Vigilance. Psychol. Rev., 62, 5, 359-368.

ELLIOT, E. (1955) Unpublished Report.

FRASER, D.C. (1950) The relation between angle of displays and performance in a prolonged visual task. Quart. J. Exp. Psychol., 1950, 2, 176-181.

FRASER, D.C. (1951) Unpublished thesis.

FRASER, D.C. (1952) A Study of Fatigue in Aircrew. Pt. I. A.P.U. Report 185, 1952. Applied Psychol. Res. Unit, Cambridge.

FRASER, D.C. (1953a) The Relation of an Environmental Variable to Performance in a Prolonged Visual Task, Quart. J. Exptl. Psychol., 5, 37-38 (1953).

FRASER, D.C. (1953b) Research on Fatigue, Discovery XIV, 2, 40-44.

FRASER, D.C. (1954) A Psychological Glossary. W. Heffer & Sons, Ltd., Cambridge.

FRASER, D.C. (1955) A Study of Fatigue in Aircrew, Pt. II. F.P.R.C. Report No. 925, Air Ministry, London.

FRASER, D.C. (1956a) Psychology in Aviation Medicine, from Handbook of Aviation Medicine, H.M.S.O., London, (To be published).

FRASER, D.C. (1956b) in National Coal Board Medical Research Memorandum No. 1, (To be published).

HEAD, H. (1926) Aphasia, Cambridge Univ. Press.

LIND, A.R. HELLON, R.F., WEINER, J.S., & JONES, R.M. (1955). Tolerance of Men to Work in Hot Saturated Environments with reference to Mines Rescue Operations, Brit. J. Indust. Med., 12, 296.

LINDSLEY, D.B. et al. (1944). Radar Operator "fatigue". <u>OSRD Report</u> No. 3334, Jan. 4, 1944.

MACKWORTH, N.H. (1950) Researches on the Measurement of Human Performance. Med.Res.Council Special Report Series, No. 268 H.M.S.O.

SIDDALL, G.J. and ANDERSON, D.M. (1955). Fatigue During Prolonged Performance on a Simple Compensating Tracking Task. Quart.J. Exptl.Psychol. VII, 4, 159-165.

WELFORD, A.T., BROWN, R.A. and GABB, J.E. (1950). Two experiments on fatigue as affecting skilled performance in civilian aircrew. Brit. J. Psychol. XL. Part 4.

WYATT, C., and LANGDON, J.N. (1932). Inspection processes in industry. <u>Rep. Indust. Health. Res. Bd.</u> No. 63, London. H.M.S.O. 1932.